ARMY RESEARCH LABORATORY

An Analysis of Optimal Training for Correlated Fading Channels Using Cutoff Rate

by Saswat Misra, Ananthram Swami, and Lang Tong

ARL-TR-3193   **April 2004**

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory

Adelphi, MD 20783-1197

An Analysis of Optimal Training for Correlated Fading Channels Using Cutoff Rate

Saswat Misra and Ananthram Swami
Computational and Information Sciences Directorate, ARL

Lang Tong
Department of Electrical Engineering, Cornell University

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
April 2004	Final	FY03 to FY04

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
An Analysis of Optimal Training for Correlated Fading Channels Using Cutoff Rate	
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
Saswat Misra, Ananthram Swami, and Lang Tong	
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
U.S. Army Research Laboratory ATTN: AMSRD-ARL-CI-CN 2800 Powder Mill Road Adelphi, MD 20783-1197	ARL-TR-3193

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
U.S. Army Research Laboratory 2800 Powder Mill Road Adelphi, MD 20783-1197	
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

This worked completed under Communications and Networks CTA

14. ABSTRACT

We consider the problem of optimal allocation of resources (power and bandwidth) between training and data for transmission over a Gauss-Markov fading channel. Inaccurate channel state information (CSI) is available at the receiver through periodic training. The transmitter is assumed to have knowledge of the channel Doppler spread. We compute the ergodic cutoff rate R_o for the temporally-correlated Rayleigh flat-fading channels with imperfect channel state information (CSI) available at the receiver. We study MMSE estimators that predict the current channel state based on: all past pilot observations, only the most recent pilot observation, and the most recent and next in the future pilot observations. We analyze the optimal training energy and periodicity for each of these estimators. We show that optimizing the energy and periodicity of training results in significant energy savings over a sensible, but unoptimized, approach. The robustness of our analysis is tested for the case where the Doppler spread is known imperfectly at the transmitter. Additionally, we consider performance under the Jakes model.

15. SUBJECT TERMS

Cutoff rate, correlated fading, Jakes Model, training, Gauss-Markov, Doppler

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Saswat Misra
a. REPORT	b. ABSTRACT	c. THIS PAGE	SAR	87	19b. TELEPHONE NUMBER (Include area code) 301-394-5640
Unclassified	Unclassified	Unclassified			

Contents

List of Figures

List of Tables

Notation and Definitions

1. $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a complex Gaussian random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and with independent real and imaginary parts, each having covariance matrix $\boldsymbol{\Sigma}$.

2. $\mathcal{E}[.]$ is the expectation operator.

3. Superscript "$*$" denotes complex conjugation.

4. Superscript "t" denotes the transpose of a matrix.

5. Superscript "H" denotes the conjugate transpose of a matrix.

6. $[k] \triangleq k \mod T$, where T is the training period.

7. $\lceil a \rceil$ denotes the ceiling of the real number a (i.e., the smallest integer larger than or equal to a.

8. $(A)_{ij}$ is the element in i^{th} row and j^{th} column ($0 \leq i, j \leq M-1$) of the $M \times M$ matrix A.

9. $\mathcal{Z}(p(\tau))$ denotes the Z-transform of the sequence $p(\tau)$. Similarly $\mathcal{Z}^{-1}(P(Z))$ denotes the inverse transform.

10. I_N is the $N \times N$ identity matrix.

11. $A \otimes B$ denotes the Kronecker product of matrices A and B.

12. $A \odot B$ denotes the Hadamard product of matrices A and B.

13. In the appendices, we often use $T' \triangleq T-1$ for brevity.

1. Introduction

Practical communication systems devote part of their power and bandwidth resources to training. Increasing the power (or bandwidth) allocated to training improves the channel estimate, but decreases the power (bandwidth) available for data transmission. We expect that if the channel varies more rapidly, the frequency of training would have to be increased, and if the channel is noisier, the training energy would have to be increased. But channel mean square error translates non-linearly into (uncoded) bit error rate (BER), and computing the impact on coded BER is difficult. To understand these fundamental tradeoffs, we use the cutoff rate as a metric.

The channel cutoff rate R_o was first proposed by Wozencraft and Jacobs (e.g., see [34]), and later popularized by Massey [23], who proposed the cutoff rate as the appropriate metric for evaluating the performance of various modulation schemes in coded systems. The cutoff rate plays a dual role: being a lower bound on the channel capacity, R_o indicates a range of rates R over which reliable communication is possible, $0 < R < R_o$, and it also gives a meaningful bound on the error performance P_e of N-length block coding (at any $R < R_o$) via the expression $P_e \leq 2^{-N(R_o-R)}$. In particular, the cutoff rate specifies the largest linear function $R_o - R$ which lower bounds the random coding exponent of [12]. Contradicting earlier opinion, it is now known that the cutoff rate can be exceeded under finite complexity constraints, with, e.g., turbo coding and iterative decoding schemes [5]. However, the cutoff rate can still be regarded as a "practical channel capacity" for simple encoding/decoding strategies. In particular, R_o has been proven to play exactly this role for the case of sequential decoding [3].

Many previous works consider the performance of coded communications systems operating over the Rayleigh fading channel under the assumption that either perfect channel state information (CSI) or no CSI is available at the receiver. In the case of no CSI, the cutoff rate [15], [28] and capacity [1], [9] have been examined.[*] With perfect CSI, the capacity has been found in [13], [7]. It is known that, with perfect CSI, the capacity is invariant with respect to the time-correlation of the channel. Conversely, with perfect CSI at the receiver, the cutoff rate is maximized for i.i.d. fading, and generally decreases as the channel correlation increases [24]. This apparent contradiction is resolved by noting that capacity considers only the horizontal axis of the random coding exponent (i.e., the maximum reliable rate), whereas the cutoff rate characterizes the vertical axis (i.e., the magnitude) of the exponent [19]. The cutoff rate with perfect CSI has been examined for i.i.d. fading [17], as well as temporally correlated fading in [21].

---

[*]In this paper, CSI refers strictly to receiver CSI.

In the intermediate case of imperfect receiver CSI, one way to characterize reliable rates, and define optimized estimation parameters, is to first fix a particular channel estimation scheme (or "front end"), and then maximize the information theoretic metric of interest (e.g., cutoff rate or mutual information) over the relevant system design parameters. This simplifying approach is used in both *[14]* and *[26]* for a training-based, minimum mean square error (MMSE) channel estimation front-end. In both cases, the authors have considered the mutual information metric.

We consider a periodic training scheme as in *[26]*. Pilot symbols are sent periodically to provide (inaccurate) estimates of the flat-fading channel coefficient. Knowledge of the channel correlation allows us to predict the fading channel between pilot symbols. We consider three MMSE channel predictors: (a) the $\mathbf{Q}_{(\mathbf{1,0})}$ estimator uses only the last pilot; (b) the $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimator uses all of the past pilots; (c) the $\mathbf{Q}_{(\mathbf{1,1})}$ estimator uses the last and next pilots (non-causal). Using the cutoff rate as a metric, we address the following issues: What is the optimal energy allocation between data and training? What is the optimal training frequency? To illustrate the approach, we focus on a first-order Gauss-Markov model that has often been used to characterize fading channels. We quantify our results in terms of an α parameter that measures how rapidly the channel is fading ($\alpha = 0$ corresponds to i.i.d. fading, and $\alpha = 1$ to a static channel). To gauge the performance of our techniques in real-world scenarios, we also consider the Jakes model, and also the impact of imperfect knowledge of the Doppler spread α at the transmitter.

Major results are as follows:

(a) Given the MMSE estimation front-end, we derive an expression for the cutoff rate for generalized binary signaling with partial CSI. For any fixed input, the cutoff rate takes on a simple closed-form expression that is amenable to analysis; it is parameterized by the mean square estimation error and the received SNR.

(b) The cutoff rate expression holds for any MMSE estimator that forms its estimate based on any subset of past and future pilots. We provide the mean square error for the three different estimators listed previously.

(c) When BPSK is used, we consider the Gauss-Markov fading channel and derive exact expressions for the optimal allocation of energy between the pilot and data transmission for the $\mathbf{Q}_{(\mathbf{1,0})}$ and $\mathbf{Q}_{(\mathbf{1,1})}$ estimators. For each of the three estimators, we find a lower bound on the optimal training frequency that is exact at high SNR.

(d) We consider the scenario where the transmitter has incorrect knowledge of the Doppler spread. We find that overestimating the Doppler spread (i.e., assuming that the channel is slower varying than it is) results in a drastic degradation of the cutoff rate. Conversely, the cutoff rate is robust to underestimate of the Doppler spread.

(e) We confirm that our analytic insights for optimal training energy and training frequency, based on the Gauss-Markov channel model, are indeed in close agreement to those provided using a numerical analysis of the more popular Jakes model.

(f) Among binary distributions, On-Off Keying (OOK) is optimal when no CSI is available, and BPSK is optimal when full CSI is present. Motivated by this fact, we derive an analytic design rule that dictates which of these two distributions to use when only partial CSI is available at the receiver.

2.   System Model

In this section, we first describe the Gauss-Markov fading channel. Next, we describe the estimation scheme and derive the variance of three different MMSE estimators of varying complexity.

2.1   Gauss-Markov Channel Model

We consider a temporally correlated Rayleigh flat-fading channel model. Typically the Jakes model [18] is used to describe the temporal correlation of the fading process. It is known that second- and third-order autoregressive (Gauss-Markov) models provide excellent fits to the Jakes model [35]. The higher-order models are not analytically tractable and so do not provide ready insights. We consider the Gauss-Markov fading channel whose correlation function is described by a first order autoregressive process. The Gauss-Markov channel has previously been used to characterize the effect of imperfect channel knowledge on the performance of decision-feedback equalization [25], mutual information [27], and minimal mean square estimator error [10] of time-correlated faded communications links. Letting the subscript k denote the time index, the observed signal at the receiver y_k' is given by

$$y_k' = \sqrt{E_k} h_k' s_k + n_k', \tag{1}$$
$$h_k' = \alpha h_{k-1}' + z_k.$$

where $h_k' \sim \mathcal{CN}\left(0, \sigma_h^2\right)$ describes Rayleigh fading, the coded input s_k is selected from a binary signal set $\mathcal{S} = \{A, B\}$ (i.e., $s_k \in \mathcal{S}$) and subject to a unit average-energy constraint: $pA^2 + (1 - p)B^2 = 1$, where p is the probability of transmitting A. The random variable $n_k' \sim \mathcal{CN}\left(0, \sigma_N^2\right)$ describes additive white Gaussian noise (AWGN), and the transmission energy used at time k is $E_k|s_k|^2$. The parameter α $(0 < \alpha < 1)$ describes the correlation between successive channel states and is related to the normalized Doppler spread of the channel. The Gauss-Markov channel is equivalently specified by its correlation function $R_A(\tau) \triangleq \frac{1}{\sigma_h^2} \mathcal{E}\left[h_k' h_{k+\tau}'^*\right] = \alpha^{|\tau|}$.

2.2 Channel Estimation

To obtain (imperfect) channel state information at the receiver, we assume that pilot (i.e., known) symbols are periodically inserted into the symbol stream. Specifically, we consider a scheme in which a single pilot symbol is inserted periodically into the symbol stream, with period T. The choice of periodic pilot placement is natural given the wide-sense stationarity of the channel. Motivation for inserting a single pilot at a time, rather than several, may be found in [2], [10], and [8].

Training is sent periodically once every T transmissions, during the time slots $k = mT$. At each time k, an estimate of the channel $\widehat{h}'_k$ is made at the receiver. Denoting the estimation error by $\widetilde{h}'_k$, the system equation can be written as

$$
\begin{aligned}
y'_k &= \sqrt{E_{\lceil k/T \rceil}}\, h'_k s_k + n'_k \\
&= \sqrt{E_{\lceil k/T \rceil}}\, \widehat{h}'_k s_k + \sqrt{E_{\lceil k/T \rceil}}\, \widetilde{h}'_k s_k + n'_k,
\end{aligned}
\tag{2}
$$

where we have assumed that the energy allocation for all *data* transmissions E_1 is constant due to practical constraints, e.g., peak-to-average-power ratio specifications and transmitter complexity (we consider variable-energy data slots in section 5.) The input s_k is now selected from a complex signal set $\mathcal{S}_{[k]} = \{A_{[k]}, B_{[k]}\}$ and subject to a unit average-energy constraint: $p_{[k]}A_{[k]}^2 + (1 - p_{[k]})B_{[k]}^2 = 1\ \forall k \neq mT$. In the training slots, we assume, without loss of estimator performance, that $\mathcal{S}_0 = \{+1\}$. Given the T-periodic nature of the estimation process, we require that codewords of length $N' = N(T-1), N \in \mathcal{Z}$ be used.

The particulars of the estimation process follow: At each time $mT + \ell$, an MMSE estimate of the channel $\widehat{h}'_{mT+\ell}$ is made at the receiver using some subset $\mathcal{N}$ of past (and possibly future) training symbol observations, so that

$$
\widehat{h}'_{mT+\ell} = \mathcal{E}[h'_{mT+\ell}|\{y'_{nT}\},\ n \in \mathcal{N} \subset \mathcal{Z}].
\tag{3}
$$

The use of an MMSE estimator implies that $\widehat{h}'_{mT+\ell} \sim \mathcal{CN}(0, \widehat{\sigma}_\ell^2)$, where $\widehat{\sigma}_\ell^2$ is the estimator variance. From orthogonality, $\widetilde{h}'_{mT+\ell} \sim \mathcal{CN}(0, \sigma_h^2 - \widehat{\sigma}_\ell^2)$. That is, $\widehat{h}'_{mT+\ell}$ and $\widetilde{h}'_{mT+\ell}$ are independent. To characterize the performance of a particular estimator, we will define the *estimator quality* ω_ℓ as

$$
\omega_\ell \triangleq \widehat{\sigma}_\ell^2 / \sigma_h^2.
\tag{4}
$$

Note that orthogonality implies that $0 \leq \omega_\ell \leq 1$. Let

$$
\kappa_0 \triangleq \sigma_h^2 E_0 / \sigma_N^2\ ; \qquad \kappa_1 \triangleq \sigma_h^2 E_1 / \sigma_N^2
$$

denote the faded pilot and data energies. We consider the following three estimators (derivations are given in appendix A with $R_h(\tau) = \alpha^{|\tau|}$):

E1. The $\mathbf{Q}_{(1,0)}$ estimator uses only the most recent pilot observation to predict the subsequent $T-1$ channel states before the next pilot, i.e., the channel estimate ℓ positions after the most recent pilot is given by $\widehat{h}'_{mT+\ell} = \mathcal{E}\left[h'_{mT+\ell}|y'_{mT}\right]$. Evaluating, we find that

$$\widehat{h}'_{mT+\ell} = \frac{\sqrt{E_0}\sigma_h^2}{E_0\sigma_h^2 + \sigma_N^2}\alpha^\ell y'_{mT},$$

$$\omega_\ell^{(1,0)} = \alpha^{2\ell}\frac{\kappa_0}{1+\kappa_0}. \tag{5}$$

E2. The $\mathbf{Q}_{(\infty,0)}$ estimator uses all past pilots to predict the current channel state, i.e., $\widehat{h}'_{mT+\ell} = \mathcal{E}\left[h'_{mT+\ell}|\left\{y'_{pT}\right\}_{p=-\infty}^m\right]$ with

$$\omega_\ell^{(\infty,0)} = \alpha^{2\ell}\left[1 - \frac{1}{\frac{1}{2}\left(1+\kappa_0\right) + \sqrt{\frac{1}{4}(1+\kappa_0)^2 + \kappa_0\frac{\alpha^{2T}}{1-\alpha^{2T}}}}\right]. \tag{6}$$

Additional details on this derivation are given in appendix B.

E3. The $\mathbf{Q}_{(1,1)}$ estimator is a non-causal smoother which uses the last and "next" pilot observations to predict the current channel state, i.e., $\widehat{h}'_{mT+\ell} = \mathcal{E}\left[h'_{mT+\ell}|y'_{mT}, y'_{(m+1)T}\right]$. Evaluating, we find that

$$\widehat{h}'_{mT+\ell} = \frac{\sigma_h^2}{\sigma_N^2}\sqrt{E_0}\left[\Gamma_{(\ell)}\ y'_{mT} + \Gamma_{(T-\ell)}\ y'_{(m+1)T}\right],$$

$$\omega_\ell^{(1,1)} = (\Gamma_{(\ell)}^2 + \Gamma_{(T-\ell)}^2)(\kappa_0^2 + \kappa_0) + 2\kappa_0^2\Gamma_{(\ell)}\Gamma_{(T-\ell)}\alpha^T, \tag{7}$$

where

$$\Gamma_{(k)} \triangleq \frac{\alpha^k(\kappa_0+1) - \alpha^T\alpha^{T-k}\kappa_0}{(\kappa_0+1)^2 - \kappa_0^2\ \alpha^{2T}}.$$

We assume that perfect interleaving is performed at the transmitter [4], and that channel estimation is done before de-interleaving at the receiver (see appendix C for a discussion of interleaving). The observation equation is now

$$y_k = \sqrt{E_k}h_k s_k + n_k,$$

where h_k is an i.i.d. sequence representing the interleaved channel, and where $n_k \sim \mathcal{CN}\left(0, \sigma_N^2\right)$ is another AWGN sequence. Writing this in terms of the channel estimate and error, we have

$$y_k = \sqrt{E_{\lceil k/T\rceil}}\ h_k s_k + n_k$$
$$= \sqrt{E_{\lceil k/T\rceil}}\ \widehat{h}_k s_k + \sqrt{E_{\lceil k/T\rceil}}\ \widetilde{h}_k s_k + n_k, \tag{8}$$

where $\widehat{h}_k$, and $\widetilde{h}_k$ are independent of each other and are independent sequences with the same marginal statistics as in equation 2. That is, interleaving does not change the

marginal statistics of the channel estimate and estimation error: $\widehat{h}_{mT+\ell} \sim \mathcal{CN}\left(0, \widehat{\sigma}_\ell^2\right)$ and $\widetilde{h}_{mT+\ell} \sim \mathcal{CN}\left(0, \sigma_h^2 - \widehat{\sigma}_\ell^2\right)$. The estimator quality is defined by equation 4 as before. Finally, we assume that codewords are decoded using the ML-detector which treats $s_{mT+\ell}$ as the channel input and the pair $(y_{mT+\ell}, \widehat{h}_{mT+\ell})$ as the channel output.

In section 3, it will be seen that the cutoff rate R_o is an (increasing) function of ω_ℓ. Therefore, it is useful to compare the estimator quality expressions in equations 5 to 7, as we have done previously in [30]: Note that the estimator quality of the $\mathbf{Q}_{(1,0)}$ and $\mathbf{Q}_{(\infty,0)}$ estimators decreases monotonically with ℓ, i.e., distance from the last pilot; the quality of the $\mathbf{Q}_{(1,1)}$ estimator is symmetric, with worst performance midway between the two pilots. It can be verified that $\omega_\ell^{(1,0)} \leq \omega_\ell^{(\infty,0)}$, that $\omega_\ell^{(1,0)} \leq \omega_\ell^{(1,1)}$, and that no similar inequality can be given between $\omega_\ell^{(\infty,0)}$ and $\omega_\ell^{(1,1)}$. In the high SNR regime ($\kappa_0 \to \infty$), $\omega_\ell^{(1,0)} \to \omega_\ell^{(\infty,0)}$. This is because the channel is learnt *perfectly* in each pilot slot, and so additional past pilots do not improve the estimator. For a rapidly fading channel, as $\alpha \to 0$, $\omega_\ell^{(1,0)} \to \omega_\ell^{(\infty,0)}$, since it is the most recent pilot that provides most of the information about the channel state. For a nearly static channel, i.e., as $\alpha \to 1$, $\omega_\ell^{(\infty,0)} > \omega_\ell^{(1,1)}$. This is because the $\mathbf{Q}_{(\infty,0)}$ estimator provides an infinite number of noisy looks at the static channel, whereas the $\mathbf{Q}_{(1,1)}$ estimator provides only two noisy looks. Further, $\omega_\ell^{(\infty,0)} < \omega_\ell^{(1,1)}$ if the channel is varying rapidly or the SNR is large: When $\alpha \to 0$, it is the closest pilots that contain channel information; the $\mathbf{Q}_{(1,1)}$ estimator provides two "close" pilots. As $\kappa_0 \to \infty$, the $\mathbf{Q}_{(\infty,0)}$ estimator converges to the $\mathbf{Q}_{(1,0)}$ estimator, which is outperformed by the $\mathbf{Q}_{(1,1)}$ estimator. Lastly, we note that all three estimators become equivalent for high-SNR static channels, i.e., as $\alpha \to 1$ *and* $\kappa_0 \to \infty$.

3. Cutoff Rate

In this section we derive the cutoff rate for general binary signaling given the estimation front-end described in section 2. The results and discussion in this section were first given (without proofs) in [28].

The cutoff rate for the system described by equation 8 is given by

$$
R_o = -\min_{Q(.)} \frac{1}{T} \log_2 \int_{\mathbf{y}} \int_{\widehat{\mathbf{h}}} \left[\sum_{\mathbf{s} \in \mathcal{S}_1 \times \ldots \times \mathcal{S}_{T-1}} Q(\mathbf{s})\sqrt{\mathbf{P}(\mathbf{y}, \widehat{\mathbf{h}} \mid \mathbf{s})} \right]^2 d\widehat{\mathbf{h}}\, d\mathbf{y}
$$

$$
= -\min_{Q(.)} \frac{1}{T} \log_2 \mathcal{E}_{\widehat{\mathbf{h}}} \left[\int_{\mathbf{y}} \sum_{\mathbf{s},\mathbf{v} \in \mathcal{S}_1 \times \ldots \times \mathcal{S}_{\mathbf{T-1}}} Q(\mathbf{s})Q(\mathbf{v}) \sqrt{P(\mathbf{y} \mid \widehat{\mathbf{h}},\, \mathbf{s})} \sqrt{P(\mathbf{y} \mid \widehat{\mathbf{h}},\, \mathbf{v})}\, d\mathbf{y} \right] \quad (9)
$$

where $\widehat{\mathbf{h}} \triangleq \left[\widehat{h}_{mT+1}, \ldots, \widehat{h}_{(m+1)T-1}\right]^t$, $\mathbf{y} \triangleq \left[y_{mT+1}, \ldots, y_{(m+1)T-1}\right]^t$, and

$\mathbf{s} \triangleq \left[s_{mT+1}, \ldots, s_{(m+1)T-1} \right]^t$ are the estimated channel, observation, and signal corresponding to the mth T-length "super-symbol." The input distribution $Q(\mathbf{s}) = \prod\limits_{\ell=1}^{T-1} Q_\ell(s_{mT+\ell})$, where $Q_k(A_k) = p_k$ and $Q_k(B_k) = 1 - p_k$.

In appendix D it is shown that the cutoff rate for generalized binary transmission becomes (setting $E_\ell = E_1$ in the appendix)

$$R_o = -\frac{1}{T} \sum_{\ell=1}^{T-1} \log_2 \min_{p_\ell A_\ell^2 + (1-p_\ell) B_\ell^2 = 1}$$
$$\left[1 + 2p_\ell (1 - p_\ell) \left\{ \frac{\sqrt{1 + \kappa_1 (1 - \omega_\ell) |A_\ell|^2} \sqrt{1 + \kappa_1 (1 - \omega_\ell) |B_\ell|^2}}{1 + \frac{1}{2} \kappa_1 (1 - \omega_\ell) (|A_\ell|^2 + |B_\ell|^2) + \frac{1}{4} \kappa_1 \omega_\ell |A_\ell - B_\ell|^2} - 1 \right\} \right], \quad (10)$$

where $\kappa_1 \triangleq \sigma_h^2 E_1 / \sigma_N^2$ was previously defined as the faded SNR during data transmissions. Next, we too consider two important special case input distributions: OOK and BPSK.

3.1 On-Off Keying

It is shown in appendix E, that when $\omega_\ell = 0$, the optimal input for the ℓth slot is a form of OOK, for which $B_\ell = 0$ and $|A_\ell|^2 = 1/p_\ell > 0$. For general ω_ℓ the OOK cutoff rate becomes

$$R_o = -\frac{1}{T} \sum_{\ell=1}^{T-1} \log_2 \min_{p_\ell A_\ell^2 = 1} \left[1 + 2p_\ell (1 - p_\ell) \left\{ \frac{\sqrt{1 + \kappa_1 (1 - \omega_\ell)/p_\ell}}{1 + \frac{1}{4} \kappa_1 (2 - \omega_\ell)/p_\ell} - 1 \right\} \right].$$

In general, it is not possible to analytically maximize equation 10 over p_ℓ, as it leads to solving a high-order polynomial that has no explicit solution as a function of κ_1 and ω_ℓ. However, it is shown in appendix E that as $\kappa_1 \to \infty$, $p_\ell^* \to 1/2$, and that as $\kappa_1 \to 0$, $p_\ell^* \to 0$ (this corresponds to no information transmission). In general $0 \le p_\ell^* \le 1/2$ (i.e., the probability of being 'OFF' $\ge 1/2$). In figure 1, we plot p_ℓ^* as a function of κ_1 for several values of ω_ℓ. We see that for moderate to large values of κ, letting $p = 1/2$ is a reasonable approximation to p^*. In the sequel, we let $p_\ell = 1/2$, for which the cutoff rate becomes

$$R_o = -\frac{1}{T} \sum_{\ell=1}^{T-1} \log_2 \left[\frac{1}{2} + \frac{1}{2} \frac{\sqrt{1 + 2\kappa_1 (1 - \omega_\ell)}}{1 + \kappa_1 (1 - \frac{\omega_\ell}{2})} \right]. \quad (11)$$

To gain insight into the behavior of equation 11, we plot the kth term in the sum above in figure 2, for several values of ω_k. As $\kappa_1 \to \infty$, the kth term approaches 1 regardless of ω_k. Therefore, as $\kappa_1 \to 1$, $R_o \to T/(T-1)$ for any value of $\{\omega_1, \ldots, \omega_{T-1}\}$.

8
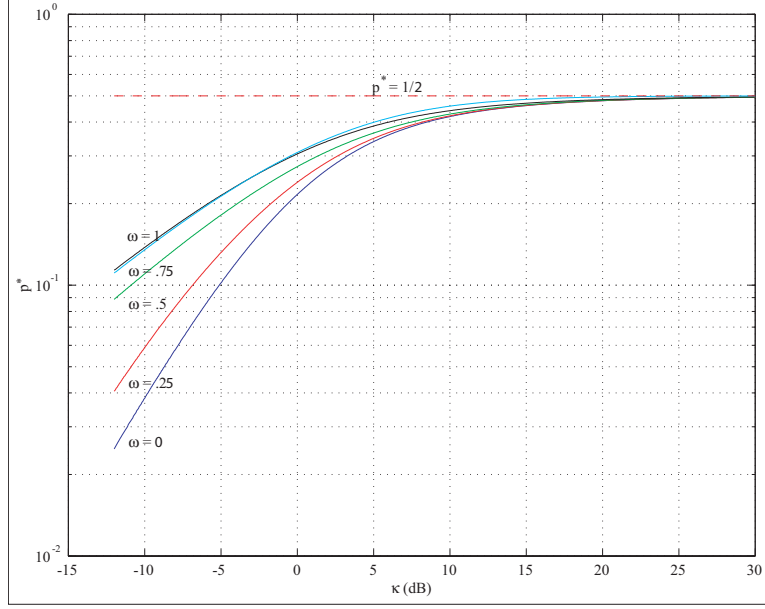
Figure 1. The optimal probability of transmitting a '1' (A), p^* vs. faded SNR κ (dB), for different values of the estimator quality ω.
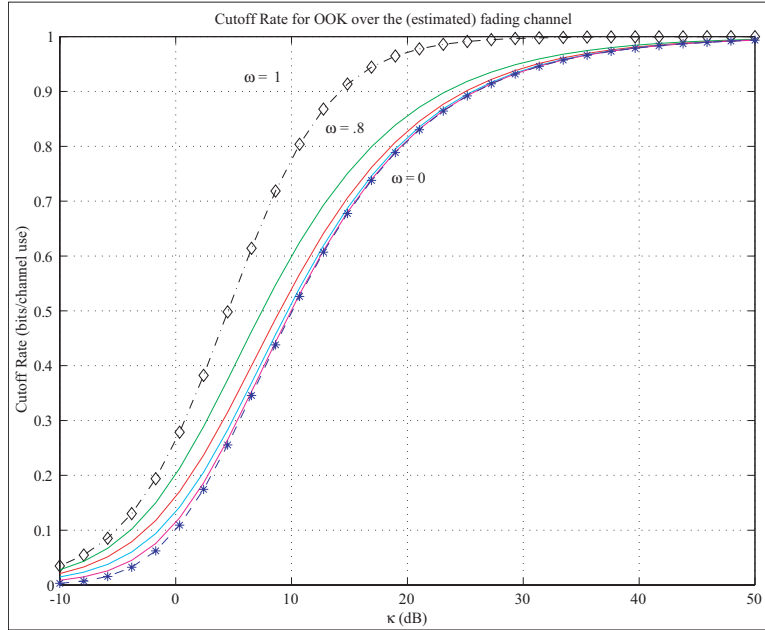
Figure 2. The OOK cutoff rate $R_{o,K}$ vs. faded SNR κ (dB), for different values of the estimator quality ω.

Since $A_\ell^2 = \frac{1}{p_\ell}$, we can equivalently repeat figure 1 in terms of the optimal OOK amplitude A_ℓ^* versus faded SNR κ_1 as shown in figure 3. Note that as κ_1 increases, the optimal OOK amplitude A_ℓ^* decreases and (correspondingly) p_ℓ^* increases. This trend was shown in [1] for the capacity metric and for no CSI ($w_\ell = 0$). From the figure, we see that when $\omega_\ell > 0$ this
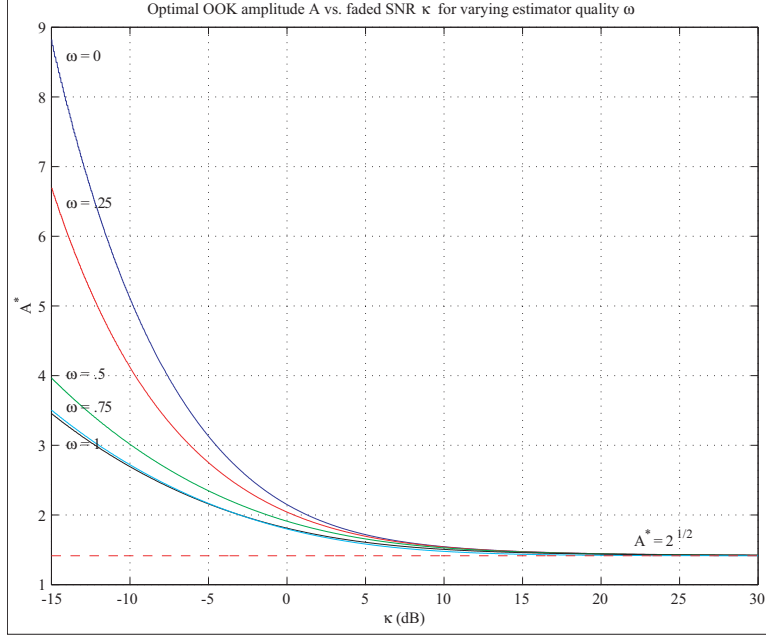
Figure 3. The optimal OOK amplitude A_ℓ^* vs. faded SNR κ_1 (dB), for different values of the estimator quality ω.

general trend is still true, and that for fixed κ_1, the optimal amplitude A_ℓ^* is a decreasing function of ω_ℓ. As κ_1 becomes large, $A_\ell^* \to \sqrt{2}$, and correspondingly, $p_\ell^* \to \frac{1}{2}$ (which is the conventional form of OOK). In section 3.3 it is noted that the value of R_o is not affected much if $p_\ell = 1/2$ is used in place of the optimal value of p_ℓ. Thus, conventional OOK can be used without sacrificing rate.

3.2 BPSK

It can be verified that, when $\omega_\ell = 1$, the optimal input in equation 10 is BPSK ($A_\ell = -B_\ell$ and $p_\ell = 1/2$). For BPSK,

$$R_o = -\frac{1}{T}\sum_{\ell=1}^{T-1}\log_2\left\{\frac{1}{2} + \frac{1}{2}\left[\frac{1 + \kappa_1(1 - \omega_\ell)}{1 + \kappa_1}\right]\right\}. \tag{12}$$

To gain insight into the behavior of equation 12, we plot the kth term in the sum above in figure 4, for several values of ω_k. We make the following observations:

1. The estimator quality ω_ℓ places an asymptotic ceiling on R_o. For large κ_1 the cutoff rate saturates to

$$R_o = -\frac{1}{T}\sum_{\ell=1}^{T-1}\log_2\left\{1 - \frac{\omega_\ell}{2}\right\}.$$

10
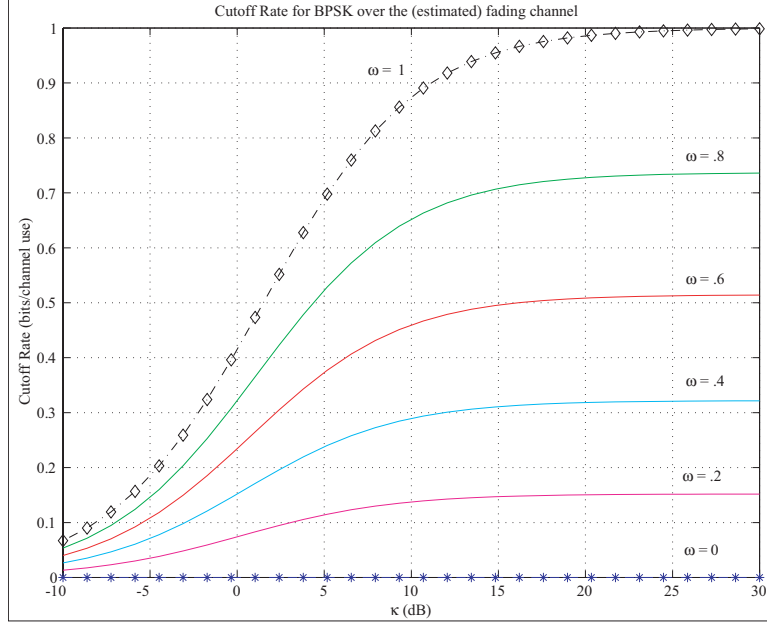
Figure 4. The BPSK cutoff rate $R_{o,B}$ vs. faded SNR κ (dB), for different values of the estimator quality ω.

2. When $\omega_\ell = 0$ (i.e., no CSI is available), information transmission is not possible. This is because the statistics of y_k at the receiver are independent of s_k; i.e.,

$$y_k|s_k = 1 \ \sim \mathcal{CN}\left(0, E_1\sigma_h^2 + \sigma_N^2\right),$$
$$y_k|s_k = -1 \ \sim \mathcal{CN}\left(0, E_1\sigma_h^2 + \sigma_N^2\right).$$

3.3 Comparisons

Here we compare the cutoff rate for BPSK to that of OOK. We start by looking at the two hypothesis for each modulation type. For BPSK, the statistics of $y_k (1 \leq k \leq T-1)$ under the two hypotheses, conditioned upon the known part of the channel $\widehat{h}_k$, are

$$y_k|\widehat{h}_k, s_k = 1 \ \sim \mathcal{CN}\left(\sqrt{E_1}\,\widehat{h}_k, E_1\widetilde{\sigma}_k^2 + \sigma_N^2\right),$$
$$y_k|\widehat{h}_k, s_k = -1 \ \sim \mathcal{CN}\left(-\sqrt{E_1}\,\widehat{h}_k, E_1\widetilde{\sigma}_k^2 + \sigma_N^2\right).$$

The ability to distinguish between the two hypotheses is only through the difference in the means, and therefore it is critical that $\omega_\ell > 0$. When $\kappa_1 \gg \frac{1}{1-\omega_\ell}$ (i.e., when $E_1\widetilde{\sigma}_k^2 \gg \sigma_N^2$), the statistics become

$$y_k|\widehat{h}_k, s_k = 1 \ \sim \mathcal{CN}\left(\sqrt{E_1}\,\widehat{h}_k, E_1\widetilde{\sigma}_k^2\right),$$
$$y_k|\widehat{h}_k, s_k = -1 \ \sim \mathcal{CN}\left(-\sqrt{E_1}\,\widehat{h}_k, E_1\widetilde{\sigma}_k^2\right).$$

11

Increasing κ_1 scales the variance and power in the mean equally, and so for large κ_1, i.e., $\kappa_1 \gg \frac{1}{1-\omega_\ell}$, performance saturates as depicted in figure 4.

Note that the OOK cutoff rate is non-zero when $\omega_\ell = 0$ (see figure 2), which is not the case for BPSK. As $\kappa_1 \to \infty$ the OOK cutoff rate approaches 1 for any value of ω_ℓ. The statistics of y_k under the two hypotheses are

$$y_k|\widehat{h}_k, s_k = 0 ~ \sim \mathcal{CN}\left(0, ~ \sigma_N^2\right)$$
$$y_k|\widehat{h}_k, s_k = \sqrt{2} ~ \sim \mathcal{CN}\left(\sqrt{2E} ~ \widehat{h}_k, 2E_1\widetilde{\sigma}_k^2 + \sigma_N^2\right).$$

The distance between the means is obviously reduced compared to that for BPSK, however the variance terms are now distinct. We expect that if the difference in the variance terms is large enough (i.e., if κ_1 is large enough), then OOK will be able to outperform BPSK despite the decreased separation between the means. Conversely, for small κ_1 (when the variance terms are nearly identical), we expect BPSK to outperform OOK. The general tradeoff is shown in figure 5 for $\omega_\ell = 0.8$. In the figure, we also plot OOK where the optimum p_ℓ is used for each value of κ_1. Note that, even for small κ_1, the difference between the two OOK curves is small; note also that there is a significant gain in using OOK instead of BPSK at moderate-to-large κ_1.

In summary, we find that OOK is preferred to BPSK when the SNR is larger, or when the estimation quality ω_ℓ is smaller. We note that BPSK is preferred to OOK for smaller κ_1 and/or for better (larger) estimation quality. This qualitative analysis is quantified in section 6.
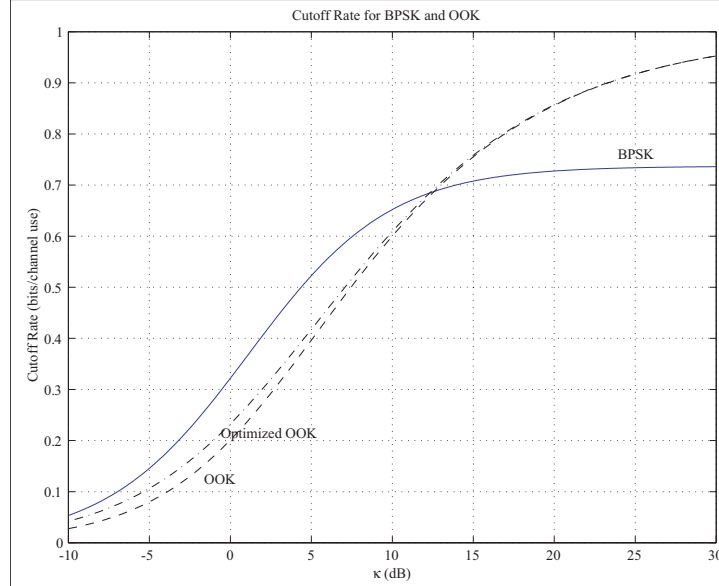
Figure 5. A comparison of the BPSK cutoff rate, the unoptimized OOK cutoff rate, and the optimized OOK cutoff rate vs. κ_1 (dB) for $\omega_\ell = 0.8$.

4. Optimized Training for BPSK

Here, we look at the optimized training parameters when the input is BPSK. We study the optimal energy allocation (κ_0^*, κ_1^*) and optimal training period T^* for the estimators E1-E3. For a meaningful analysis, we impose an average energy constraint, $\kappa_0 + (T-1)\kappa_1 = \kappa_{\mathrm{av}}T = \kappa_{\mathrm{tot}}$. First, we find (κ_0^*, κ_1^*) for a fixed T. Then, we consider the optimal value of T. We consider the case of variable energy data slots in section 5. Many of the results in sections 4.1 through 4.3 were first given in *[30]*, where proofs were omitted due to space constraints.

4.1 Optimal Energy Allocation

For the $\mathbf{Q}_{(\mathbf{1},\mathbf{0})}$ estimator, κ_1^* (from which κ_0^* follows readily) is shown in appendix F to be

$$\kappa_1^* = \Gamma - \sqrt{\Gamma^2 - \frac{\kappa_{\mathrm{av}}T}{T-1}\Gamma}, \quad \Gamma = \frac{\kappa_{\mathrm{av}}T+1}{T-2}, \tag{13}$$

for $T > 2$. For $T = 2$, $\kappa_1^* = \kappa_0^* = \kappa_{\mathrm{av}}$. Note that κ_1^* does not depend on α; however $R_o(\kappa_1^*)$ does. In the low energy regime ($\kappa_{\mathrm{tot}} \to 0$), equation 13 leads to $\kappa_0^* = (\kappa_{\mathrm{av}/2})T$, i.e., half of the available energy should be allocated to the pilot symbol. The 50-percent training paradigm has also been reported in *[14]* for a different channel model, metric, and assumptions. In the high energy regime, $(\kappa_{\mathrm{tot}} \to \infty)$, we find that $\kappa_0^* = \kappa_{\mathrm{av}}T\left[\frac{\sqrt{T-1}-1}{T-2}\right]$. For large T, the energy allocated to the training symbol increases as $\kappa_{\mathrm{av}}\sqrt{T}$; a similar result was reported in *[14]*.

For the $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimator, κ_0^* is shown in appendix F to be given implicitly by (for $T > 2$)

$$\kappa_0^* = \arg \max_{0 \le \kappa_0 \le \kappa_{\mathrm{av}} T} \left[\frac{\kappa_{\mathrm{av}}T - \kappa_0}{\kappa_{\mathrm{av}}T - \kappa_0 + (T-1)}\right] \times$$
$$\frac{\kappa_0 - 1 + \sqrt{(1+\kappa_0)^2 + 4\kappa_0\frac{\alpha^{2T}}{1-\alpha^{2T}}}}{\kappa_0 + 1 + \sqrt{(1+\kappa_0)^2 + 4\kappa_0\frac{\alpha^{2T}}{1-\alpha^{2T}}}} \tag{14}$$

This implicit solution provides useful insights. In the low energy regime, equation 14 states that $\kappa_0^* = (\kappa_{\mathrm{av}}/2)T$. In the high energy regime, $\kappa_0^* = \kappa_{\mathrm{av}}T\left[\frac{\sqrt{T-1}-1}{T-2}\right]$. In these two limiting cases, the $\mathbf{Q}_{(\infty,\mathbf{0})}$ and $\mathbf{Q}_{(\mathbf{1},\mathbf{0})}$ estimators have the same optimal energy allocation, which is independent of α. In general, κ_0^* decreases as α (which is a measure of channel predictability) increases: As $\alpha \to 1$, $\kappa_0^* \to 0$. This is because, the $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimator provides us with an infinite number of (noisy) observations of the nearly time-invariant channel. Each observation requires only a minuscule amount of energy, in order to make use of the

13

infinite diversity gain. As $\alpha \to 0$, κ_0^* converges to the κ_0^* of equation 13: for a rapidly fading channel, the most recent pilot provides all the information about the channel. In general, it is shown in appendix F that $\kappa_{0,(\infty,0)}^* \leq \kappa_{0,(1,0)}^*$; the estimator of higher quality requires less training energy.

The energy allocation rules of equations 13 and 14 can be extended in a straightforward way to any causal estimator (i.e., any estimator of the form given in equation 3, where $\max\{\mathcal{N}\} \leq m$). For any causal estimator with estimator quality ω_ℓ, κ_0^* is given implicitly in terms of the estimation quality in the pilot slot ω_0 as follows:

$$\kappa_0^* = \arg \max_{0 \leq \kappa_0 \leq \kappa_{\mathrm{av}} T} \left[\frac{\kappa_{\mathrm{av}} T - \kappa_0}{\kappa_{\mathrm{av}} T - \kappa_0 + (T-1)} \ \omega_0 \left(\kappa_0, \alpha, T, \mathcal{N} \right) \right]$$

where the notation emphasizes that ω_0 is a function of κ_0, α, T, and $\mathcal{N}$. The proof follows easily using the methodology for the $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimator, and is a consequence of the fact that, for any causal estimator, $\omega_\ell = \alpha^{2\ell} \omega_0$.

Finally, we give κ_0^* for the $\omega_\ell^{(1,1)}$ estimator. For simplicity, we consider the case where $T > 2$. Because no closed form expression for κ_0 exists in general, we will focus on the low energy, high energy, rapidly fading ($\alpha \ll 1$), and slowly fading ($\alpha \to 1$) regimes. The optimal training energy κ_0^* in the low SNR regime can again be shown to be given by $\kappa_0^* = (\kappa_{\mathrm{av}}/2)T$. In the high SNR regime, it is shown in appendix F that $\kappa_0^* \approx \kappa_{\mathrm{av}} \sqrt{T}$ with equality when T becomes large. For rapid fading $\alpha \ll 1$, it can be seen that the optimal training energy converges to that of the $\mathbf{Q}_{(\mathbf{1},\mathbf{0})}$ estimator, i.e., $\kappa_{0,(1,1)}^* \to \kappa_{0,(1,0)}^*$ as is to be expected. For slow fading $\alpha \approx 1$, it can be shown that $\kappa_{0,(1,1)}^* = -\nu + \sqrt{\nu^2 + \nu \kappa_{\mathrm{tot}}}$ with $\nu = \frac{\kappa_{\mathrm{tot}} + T - 1}{2T - 3}$. In figure 15, we plot the optimal training energy for the limiting cases of rapid fading and slow fading for $T = 8$. The slowly fading channel makes use of about 10-percent more training energy compared to the fast fading channel. This percentage decreases for smaller values of κ_{tot}. We note that $\kappa_{0,(1,1)}^* \leq \kappa_{0,(1,0)}^*$; i.e., the better estimator requires less training energy. The optimal training energy for each of the estimators is summarized in table 1.

4.2   Optimal Training Period

The preceding analysis gives insights into the optimal energy allocation (κ_0^*, κ_1^*) for a fixed training period T. Below, we consider the optimal training period T^* for each estimator.

First we consider the causal estimators. It is shown in appendix G that for all casual estimators, a lower bound on the optimal value of the training period T_B can be found by considering the high energy regime ($\kappa_{\mathrm{av}} \to \infty$) (this is equivalent to assuming that the channel is known *perfectly* in the relevant training slots). Furthermore, the lower bound T_B is exact at high SNR, and is the same for all casual estimators, depends only on α and

Table 1. The optimal training energy κ_0^* for the $\mathbf{Q_{(1,0)}}$, $\mathbf{Q_{(\infty,0)}}$ and $\mathbf{Q_{(1,1)}}$ estimators for high SNR ($\kappa_{av} \to \infty$), low SNR ($\kappa_{av} \to 0$), rapid fading ($\alpha \ll 1$), and slow fading ($\alpha \approx 1$).

	$\mathbf{Q_{(1,0)}}$	$\mathbf{Q_{(\infty,0)}}$	$\mathbf{Q_{(1,1)}}$
κ_0^*	see (13)	see (14)	= N/A
$\kappa_{av} \to 0$	$\frac{\kappa_{av}T}{2}$	$\frac{\kappa_{av}T}{2}$	$\frac{\kappa_{av}T}{2}$
$\kappa_{av} \to \infty$	$\approx \kappa_{av}\sqrt{T}$	$\approx \kappa_{av}\sqrt{T}$	$\approx \kappa_{av}\sqrt{T}$
$\alpha \to 0$	no change	converges to $\kappa_{0,(1,0)}^*$	converges to $\kappa_{0,(1,0)}^*$
$\alpha \to 1$	no change	0	$-\nu + \sqrt{\nu^2 + \nu\kappa_{tot}}; \nu = \frac{\kappa_{tot}+T-1}{2T-3}$

is given by:

$$T_B = \arg\min_T \prod_{\ell=1}^{T-1} \left[1 - \frac{\alpha^{2(nT+\ell)}}{2} \right]^{1/T} \tag{15}$$

where n, $0 \le n < \infty$, is the number of pilots between the most recent pilot and the last pilot used. For any practical scheme, $n = 0$. For the $\mathbf{Q_{(1,0)}}$ and $\mathbf{Q_{(\infty,0)}}$ estimators, $n = 0$, and the bound becomes

$$T_{B,(1,0)} = T_{B,(\infty,0)} = \arg\min_T \prod_{\ell=1}^{T-1} \left[1 - \frac{\alpha^{2\ell}}{2} \right]^{1/T}. \tag{16}$$

Although the high-SNR bounds are equal, we find that, in general, $T_{(1,0)}^* \ge T_{(\infty,0)}^*$ for the equal energy case ($\kappa_0 = \kappa_1 = \kappa_{av}$). Furthermore, it is shown in appendix G that $T_{(1,0)}^* \ge T_{(\infty,0)}^*$ for any channel with monotonically decreasing correlation function $R_h(\tau)$.[†] We see that the better estimation scheme requires *more frequent* pilots. This was noted heuristically in *[26]*. Here, it has been proven analytically.

For the $\mathbf{Q_{(1,1)}}$ estimator, the lower bound is

$$T_{B,(1,1)} = \arg\min_T \prod_{\ell=1}^{T-1} \left[\frac{1 - \frac{1}{2}\left(\alpha^{2\ell} + \alpha^{2(T-\ell)}\right)}{1 - \alpha^{2T}} \right]^{1/T}. \tag{17}$$

We find that $T_{B,(1,1)} \ge T_{B,(\infty,0)} = T_{B,(1,0)}$ and that $T_{(1,0)}^* \le T_{(1,1)}^*$.[†] Unlike the $\mathbf{Q_{(1,0)}}$ vs. $\mathbf{Q_{(\infty,0)}}$ case, the better estimation scheme requires *less* frequent training. This apparent conflict is resolved by noting that the training period is not determined by the quality ω_ℓ of the estimation scheme, but rather, by how quickly ω_ℓ "falls off" as ℓ is increased. Table 2 compares T_B to T^* for each estimator, for several values of κ_{av} and α. In general, the bound is accurate for smaller values of α and larger values of κ_{av}. From the table, when

---

[†]This can be proven for the equal energy case. We conjecture that it is also true when we optimize over the training and data energy. The conjecture is also supported by numerical evidence.

Table 2. Comparison of the optimal training period $T^*_{(.,.)}$ to the lower bound $T_{B,(.,.)}$ for several different values of α and κ_{av}.

	$T^*_{(1,0)}$	$T^*_{(\infty,0)}$	$\dfrac{T_{B,(\infty,0)}}{T_{B,(1,0)}}$	$T^*_{(1,1)}$	$T_{B,(1,1)}$
$\alpha = 0.80$					
$\kappa_{\text{av}} = 1$	3	3	**3**	4	**4**
$\kappa_{\text{av}} = 10$	3	3	**3**	4	**4**
$\kappa_{\text{av}} = 100$	3	3	**3**	4	**4**
$\alpha = 0.95$					
$\kappa_{\text{av}} = 1$	8	5	**5**	10	**7**
$\kappa_{\text{av}} = 10$	5	5	**5**	7	**7**
$\kappa_{\text{av}} = 100$	5	5	**5**	7	**7**
$\alpha = 0.99$					
$\kappa_{\text{av}} = 1$	20	11	**9**	29	**15**
$\kappa_{\text{av}} = 10$	11	10	**9**	17	**15**
$\kappa_{\text{av}} = 100$	9	9	**9**	15	**15**

$\alpha = 0.80$, the lower bound is exact for all $\kappa_{\text{av}} \geq 1$. For $\alpha = 0.95$, the lower bound is exact for $\kappa_{\text{av}} \geq 10$. For $\alpha = 0.99$, the lower bound is tight for $\kappa_{\text{av}} \geq 10$, and exact for $\kappa_{\text{av}} \geq 100$.

4.3 Cutoff Rate with Optimized Training

First, we analyze how using each of the three proposed estimators affects the *unoptimized* cutoff rate; see figure 6. We use the same energy κ_{av} in each transmission slot, and pick a fixed, but reasonable, value for the training period ($T = 10$) based on table 2, with $\alpha = 0.99$. At low SNR, there is a 3 dB gain in the cutoff rate for the $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimator over the $\mathbf{Q}_{(\mathbf{1,0})}$ estimator. As expected, this gain diminishes as SNR is increased. Also as expected, the $\mathbf{Q}_{(\mathbf{1,1})}$ estimator outperforms the other two at high SNR. The gain in using the $\mathbf{Q}_{(\mathbf{1,1})}$ estimator over the other two is as much as 3 dB at low to moderate SNR. As SNR $\to \infty$, the cutoff rate for the $\mathbf{Q}_{(\mathbf{1,1})}$ estimator saturates to 0.8514; a value that exceeds the saturation cutoff rate of 0.7829 for either of the other two estimators.

Next, we analyze how using each of three proposed estimators affects the *optimized* cutoff rate. In figure 7, we plot the cutoff rate, optimized over the energy allocation (κ_0, κ_1) *and* training period T. Note that optimizing the cutoff rate effectively narrows the gain of the more complicated estimators over the $\mathbf{Q}_{(\mathbf{1,0})}$ estimator. In particular, the $\mathbf{Q}_{(\mathbf{1,0})}$ and $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimators perform within 1 dB of each other. At high SNR, the optimized energy allocation provides no gain; even a "sloppy" energy allocation will allow the cutoff rate to
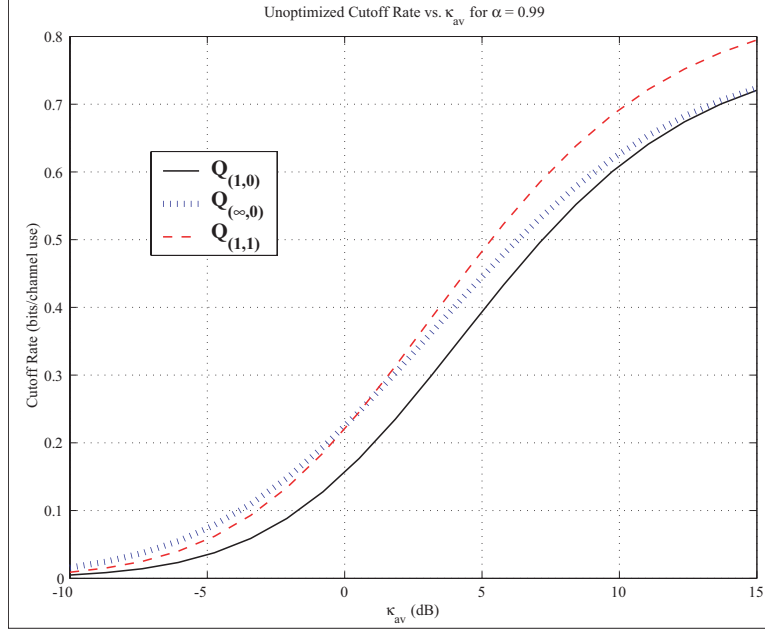
Figure 6. The cutoff rate for three different estimators for $T = 10$ for equi-energy transmission slots, and for $\alpha = 0.99$.
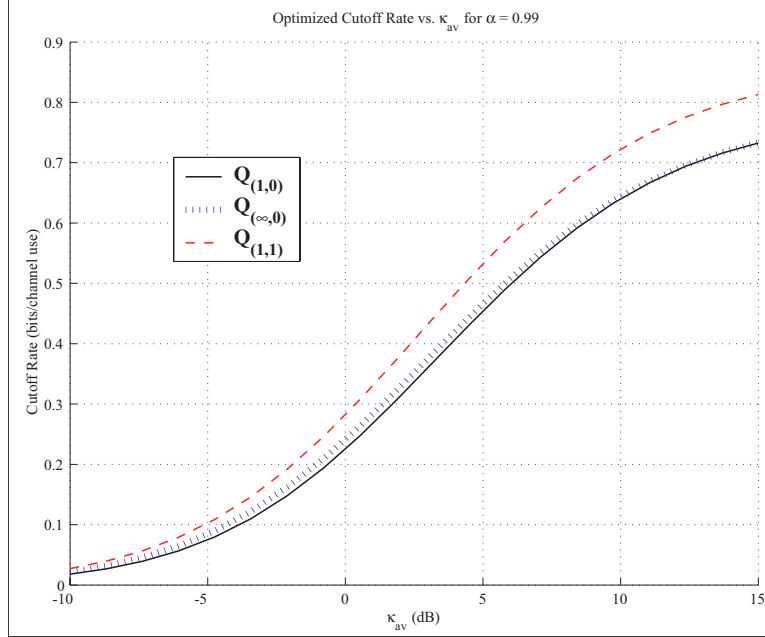
Figure 7. The cutoff rate for three different training estimators, optimized over the training period T and energy κ_0 for $\alpha = 0.99$.

saturate to its maximum value. A poor choice of T will result in a large loss of the cutoff rate relative to the optimized value, even at high SNR.

Next, we fix the particular estimator used, and by comparing figures 6 and 7, determine the gain in cutoff rate attained by using optimized training parameters in place of the unoptimized (but reasonable) parameters. For the $\mathbf{Q_{(1,0)}}$ estimator, the gain is typically between $3 \sim 4$ dB at low SNR. For the $\mathbf{Q_{(1,1)}}$ estimator, the gain is typically $2 \sim 3$ dB. For the $\mathbf{Q_{(\infty,0)}}$ estimator, the gain is typically ~ 0.5 dB. Note that the gain in using optimized parameters diminishes as the estimation scheme uses more pilot symbols. This suggests that, *as more pilot observations are exploited, the less the cutoff rate benefits from an optimized energy allocation*. We emphasize that the gain in the optimized cutoff rate would have been even more dramatic, had a "poor" value of T been chosen.

4.4 Mismatched Doppler Spread

Previously, we have assumed that the Doppler spread α is known perfectly at the transmitter. Here, we study how the cutoff rate is impacted when the transmitter has inaccurate knowledge of the Doppler spread. Given imperfect knowledge of the Doppler spread, the transmitter will, in general, incorrectly determine the energy allocation and the training frequency. In the following analysis, we assume that the receiver has perfect knowledge of the Doppler spread, and therefore, that channel prediction is carried out without error. Let $\widetilde{\alpha} = (1 + \delta)\alpha$ denote the transmitter's assumed value of the Doppler, so that δ denotes the relative error. Note that, if the transmitter adapts its rate based on its perceived value of α, it will transmit at rates larger than the cutoff rate if $\delta > 0$. This cannot be allowed, for these rates may not be supported by the channel. Hence, in this section only, we assume that the transmitter uses a *fixed transmission rate* sufficiently smaller than the cutoff rate of the channel. The cutoff rate discussed in this section should be interpreted as a bound on the probability of N-length block decoding error, given by the expression $P_e \leq 2^{-N(R_o - R)}$ for transmission rates $R < R_o$.

For the $\mathbf{Q_{(1,0)}}$ estimator, κ_0^* does not explicitly depend on α. However, incorrect knowledge of α at the transmitter will result in an incorrect assignment of T. If T is assigned incorrectly, then energy will also be allocated sub-optimally. In figure 8, we plot the normalized cutoff rate $R_o(\widetilde{\alpha})/R_o(\alpha)$ for $-0.4 \leq \delta \leq 0.5$ (corresponding to $0.57 \leq \widetilde{\alpha} \leq 0.998$) for $\alpha = 0.95$ and $\kappa_{\mathrm{av}} = 100$. Also included on the figure is the training period selected by the transmitter $\widetilde{T}$ as a function of δ. As expected, the normalized cutoff rate ≈ 1 when $\widetilde{\alpha} \approx 0.95$. From the figure, the degradation in the cutoff rate is less than 25-percent even when α is underestimated by 40-percent. When α is underestimated by up to 5-percent, there is virtually no loss in the cutoff rate. Conversely, if α is instead *overestimated* by 5-percent, there is a drastic loss in the cutoff rate, more than 35-percent. We see that it is better to underestimate α rather than to overestimate it. This is because, T^* changes more rapidly as α is increased than when it is decreased (e.g., see table 2). When α is overestimated, $\widetilde{T}$ deviates quickly from T^*. When α is underestimated, $\widetilde{T}$ deviates less quickly from T^*. This behavior is emphasized as α itself becomes large. Next,

Figure 8. The normalized cutoff rate $R_o(\widetilde{\alpha})/R_o(\alpha)$ and corresponding training period $\widetilde{T}$ for Doppler mismatch in the range $-0.4 \leq \delta \leq 0.5$, when $\alpha = 0.95$ and $\kappa_{\mathrm{av}} = 100$ for the $\mathbf{Q}_{(\mathbf{1},\mathbf{0})}$ estimator.

we make two points from figure 8: (1) $R_o(\widetilde{\alpha})$ corresponds to the perfect-α cutoff rate, if $\widetilde{T}$ were used *in place* of T^*, (2) $R_o(\widetilde{\alpha})$ changes in "discrete" steps (corresponding to changes in $\widetilde{T}$ as $\widetilde{\alpha}$ is varied. These last two properties are properties of the $\mathbf{Q}_{(\mathbf{1},\mathbf{0})}$ estimator, and are a consequence of the fact that $\kappa^*_{(1,0)}$ does not depend explicitly on α. For the other estimators, κ^*_0 depends explicitly on α and so the cutoff rate will decline continuously as $\widetilde{\alpha}$ deviates from α (as will be seen in the sequel). We note that the same general trend holds as κ_{av} is varied, as can be seen in figures 9 and 10 where we repeat the same analysis, this time for $\kappa_{\mathrm{av}} = 10$ and $\kappa_{\mathrm{av}} = 0.1$ respectively. Considering figure 10, we see that the disparity is even larger at small SNR; overestimating α by 5-percent degrades the cutoff rate by about 60-percent, whereas underestimating α by the same percentage results in a loss of less than 10-percent. We observe that, for a fixed value of δ, the normalized cutoff rate decreases with decreasing SNR.

Here, we repeat the preceding analysis for the $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimator. In figure 11, we plot the normalized cutoff rate under the same parameters as in the previous case, for $\kappa_{\mathrm{av}} = 10$. Again, we note that it is better to underestimate α rather than to overestimate it. We see that in practice the normalized cutoff rate still changes in steps that are nearly discrete (which correspond to incorrect assignments of the training period at the transmitter) for the $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimator. This implies that the degradation in the cutoff rate incurred when using a mismatched value of the Doppler spread is primarily through an incorrect assignment of T at the transmitter, and is only slightly affected by the incorrect allocation of energy (κ_0, κ_1). The normalized cutoff rate is slightly larger for the $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimator

than for the $\mathbf{Q_{(1,0)}}$ estimator, as seen by comparing figures 9 and 11. This means that the $\mathbf{Q_{(\infty,0)}}$ estimator is less sensitive to a mismatched Doppler parameter than is the $\mathbf{Q_{(1,0)}}$ estimator. We will say more on this point in what follows.
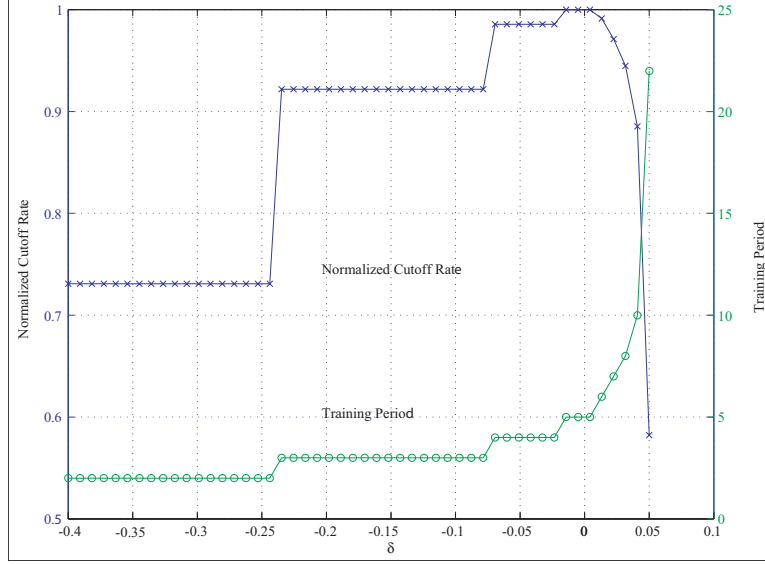
Figure 9. The normalized cutoff rate $R_o(\widetilde{\alpha})/R_o(\alpha)$ and corresponding training period $\widetilde{T}$ for Doppler mismatch in the range $-0.4 \leq \delta \leq 0.5$, when $\alpha = 0.95$ and $\kappa_{\mathrm{av}} = 10$ for the $\mathbf{Q_{(1,0)}}$ estimator.
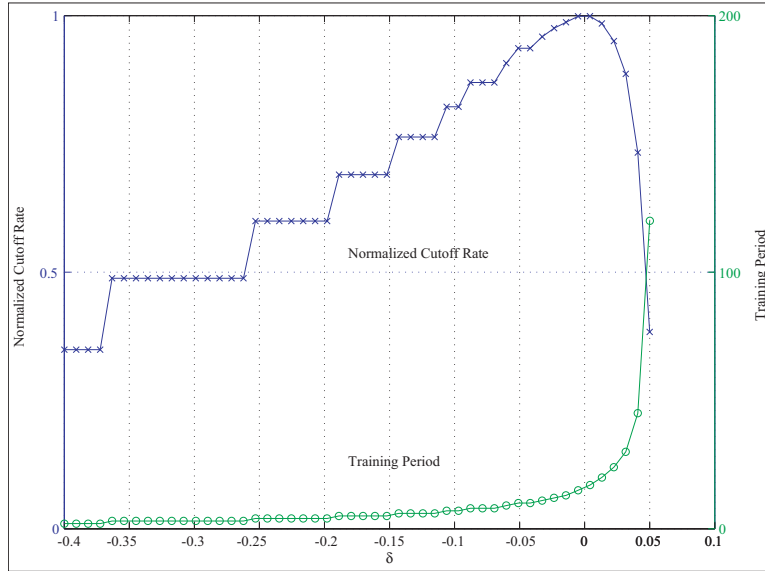
Figure 10. The normalized cutoff rate $R_o(\widetilde{\alpha})/R_o(\alpha)$ and corresponding training period $\widetilde{T}$ for Doppler mismatch in the range $-0.4 \leq \delta \leq 0.5$, when $\alpha = 0.95$ and $\kappa_{\mathrm{av}} = 0.1$ for the $\mathbf{Q_{(1,0)}}$ estimator.
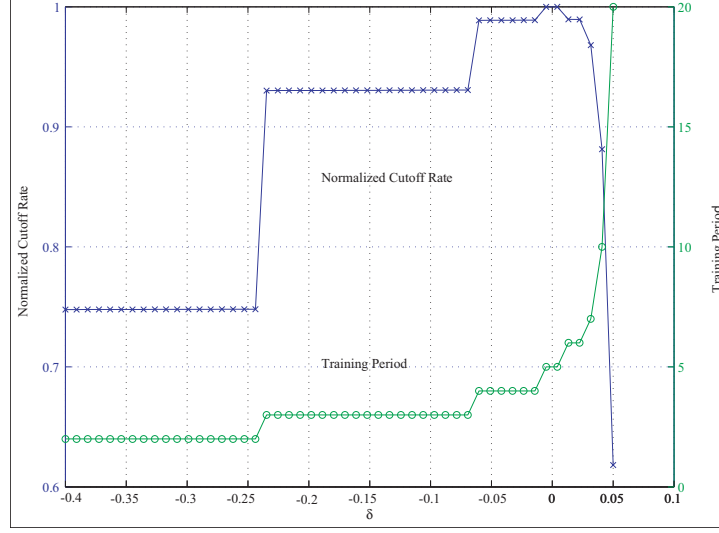
Figure 11. The normalized cutoff rate $R_o(\widetilde{\alpha})/R_o(\alpha)$ and corresponding training period $\widetilde{T}$ for Doppler mismatch in the range $-0.4 \leq \delta \leq 0.5$, when $\alpha = 0.95$ and $\kappa_{\mathrm{av}} = 10$ for the $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimator.

Next, figure 12 considers the $\mathbf{Q}_{(\mathbf{1},\mathbf{1})}$ estimator for the same parameters as in the previous two cases with $\kappa_{\mathrm{av}} = 10$. Again, we notice that it is better to underestimate α than to overestimate it and that the normalized cutoff rate changes in what are effectively discrete steps. In figure 13 we superimpose the normalized cutoff rates for all three estimators when $\kappa_{\mathrm{av}} = 10$; figure 14 shows the corresponding training period determined by the transmitter $\widetilde{T}$. We note that in general, *any estimator may be the most (or least) sensitive to a Doppler mismatch*, depending on the particular parameters chosen.

In general (i.e., over a wide variety of simulation parameters) we have observed the following trends: (1) We continue to find that it is better to underestimate α than to overestimate it. (2) We find that the normalized cutoff rate $R_o(\widetilde{\alpha})/R_o(\alpha)$ always changes in (nearly) discrete steps, for each of the three estimators. This implies that when using a mismatched value of α at the transmitter, the degradation is due primarily to choosing an incorrect value of T and that the subsequent misallocation of energy has an insignificant effect on the cutoff rate. The issue of how much the cutoff rate is degraded due to mismatched α consists of two parts: (1) how quickly does $\widetilde{T}$ change with $\widetilde{\alpha}/\alpha$? (2) how quickly does the cutoff rate degrade with $\widetilde{T}$ for a given estimator? These two issues must be considered jointly. For example, consider the initial set of parameters (where $\kappa_{\mathrm{av}} = 10$). From table 2 or figure 14, we see that $T^*_{(1,1)}$ changes most rapidly with α and is followed by $T^*_{(1,0)}$ and lastly by $T^*_{(\infty,0)}$. However, from figure 13, it is evident that the cutoff rate of the $\mathbf{Q}_{(\mathbf{1},\mathbf{1})}$ estimator changes *least* rapidly with $\widetilde{T}$ over a wide range of δ (e.g., the range $-0.16 \leq \delta \leq 0$), so that the $\mathbf{Q}_{(\mathbf{1},\mathbf{1})}$ estimator is the least sensitive to Doppler mismatch over this range.
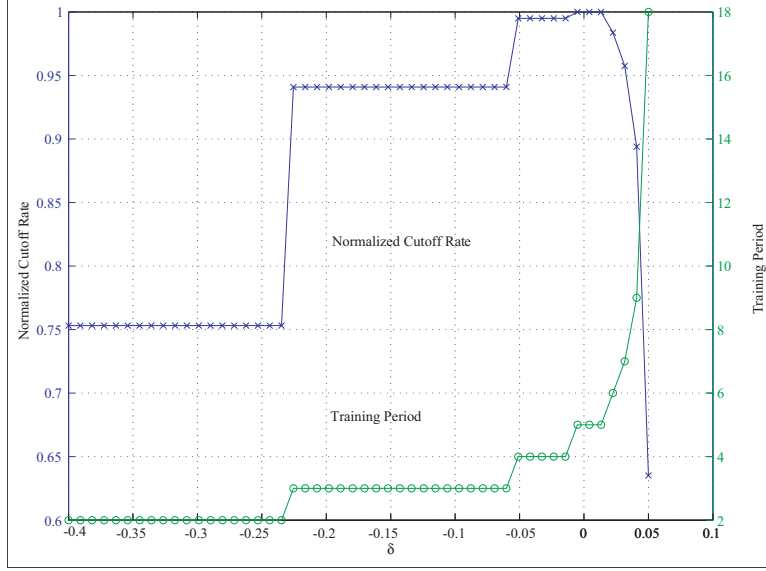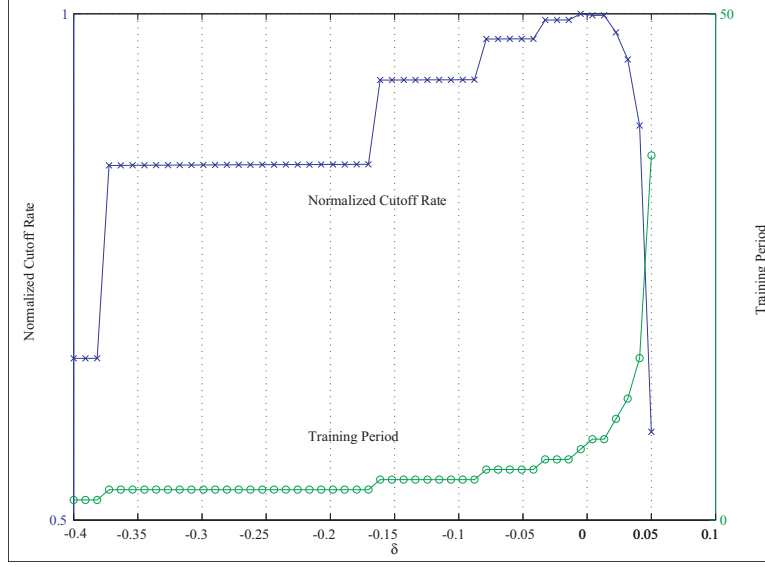
Figure 12. The normalized cutoff rate $R_o(\widetilde{\alpha})/R_o(\alpha)$ and corresponding training period $\widetilde{T}$ for Doppler mismatch in the range $-0.4 \leq \delta \leq 0.5$, when $\alpha = 0.95$ and $\kappa_{\mathrm{av}} = 10$ for the $\mathbf{Q_{(1,1)}}$ estimator.
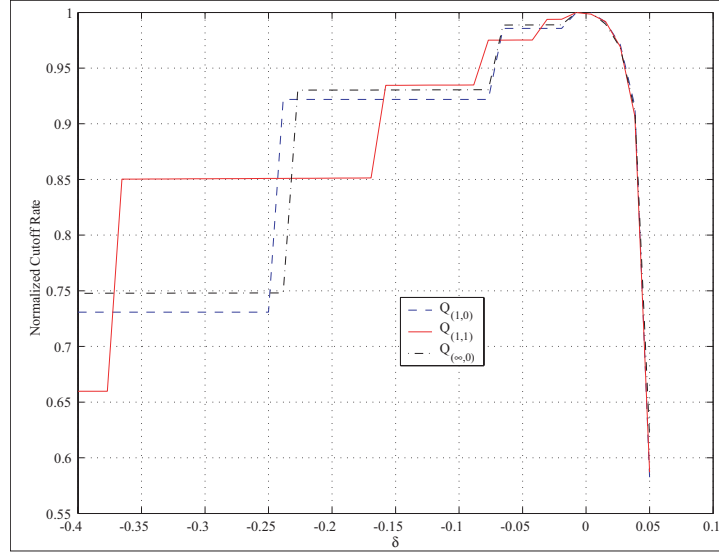
Figure 13. The normalized cutoff rate $R_o(\widetilde{\alpha})/R_o(\alpha)$ for Doppler mismatch in the range $-0.4 \leq \delta \leq 0.5$, when $\alpha = 0.95$ and $\kappa_{\mathrm{av}} = 10$ for all estimators.
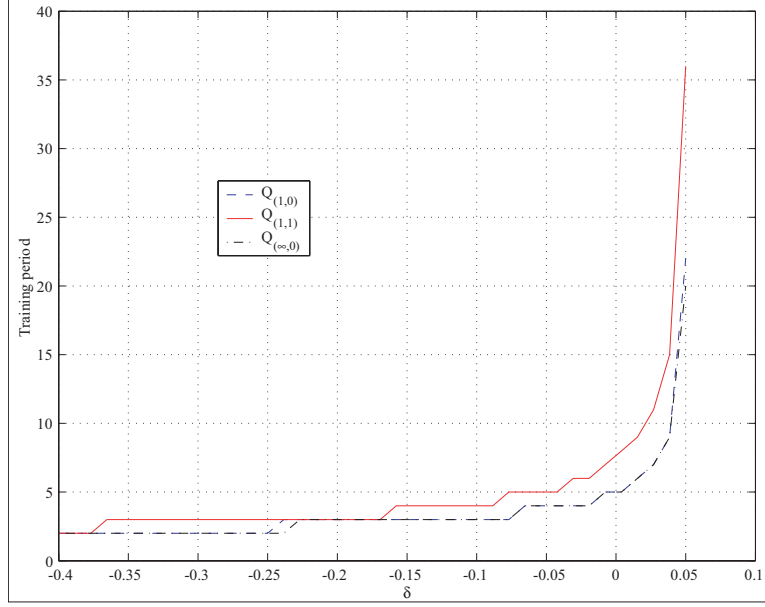
Figure 14. The training period determined by the transmitter $\widetilde{T}$ for Doppler mismatch in the range $-0.4 \leq \delta \leq 0.5$, when $\alpha = 0.95$ and $\kappa_{\mathrm{av}} = 10$ for all estimators.
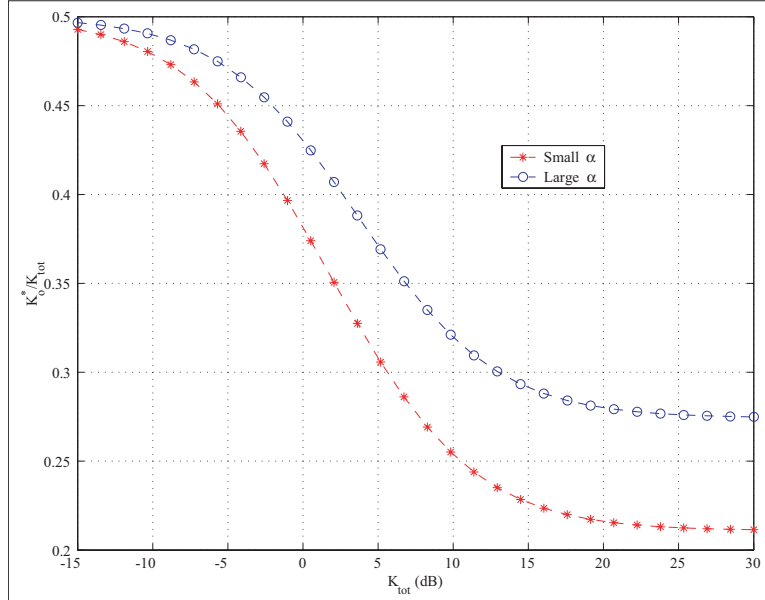
Figure 15. The optimal percentage of training energy for the $\mathbf{Q}_{(\mathbf{1},\mathbf{1})}$ estimator, $\kappa_{0,(1,1)}/\kappa_{\mathrm{tot}}$, for the limiting cases of a rapidly varying ($\alpha \ll 1$) and slowly varying ($\alpha \approx 1$) channel for $T = 8$.

5.   Variable Energy Data Slots

Here, we generalize the energy allocation problem of section 4 to the case where the kth sub-channel is allocated an arbitrary energy E_ℓ. We impose a total energy constraint $\sum_{\ell=0}^{T-1} \kappa_\ell \leq \kappa_{\text{tot}}$ (where $\kappa_\ell \geq 0$), and seek to find the energy allocation $\boldsymbol{\kappa} = \{\kappa_0, \ldots, \kappa_{T-1}\}$ that maximizes the cutoff rate. We treat the training period T as a fixed parameter, and consider the $\mathbf{Q_{(1,0)}}$ estimator. We first presented the results in sections 5.1 through 5.3 below in [28], where proofs were omitted due to space limitations.

For variable energy data slots, the observation equation in equation 8 now becomes

$$y_k = \sqrt{E_{[k]}}\, \widehat{h}_k s_k + \sqrt{E_{[k]}}\, \widetilde{h}_k s_k + n_k,$$

and the corresponding cutoff rate is (see appendix D.)

$$R_o = -\frac{1}{T} \sum_{\ell=1}^{T-1} \log_2 \left\{ \frac{1}{2} + \frac{1}{2} \left[\frac{1 + \kappa_\ell (1 - \omega_\ell)}{1 + \kappa_\ell} \right] \right\}. \tag{18}$$

In seeking the optimal energy allocation, our intuition from water-filling over parallel AWGN channels applies, as interleaving removes the correlation between the T sub-channels with respect to coding (the correlation is exploited instead in the estimator design). Water-filling predicts that more energy will be allocated to less noisy channels, and that channels with noise-levels above a threshold will not be used at all. We will see that both of these ideas are preserved.

5.1   Substitution Function

Optimization of R_o over $\boldsymbol{\kappa}$ does not lead to a closed form solution for the optimal energy allocation. Hence, we propose an approximate solution based on optimizing the substitution function

$$\widetilde{R}_o \triangleq \sum_{\ell=1}^{T-1} \alpha^{2\ell} \frac{\kappa_\ell \kappa_0}{(1 + \kappa_0)(1 + \kappa_\ell)} \tag{19}$$

over $\boldsymbol{\kappa}$. We will denote the optimizer of the substitution function by $\widetilde{\boldsymbol{\kappa}}^*$. Let $\boldsymbol{\kappa}^*$ be the optimal energy-vector for R_o in equation 18. We claim that $\widetilde{\boldsymbol{\kappa}}^* \approx \boldsymbol{\kappa}^*$ for the following reasons (proofs and further details are given in appendix H):

A1. The approximate solution is *exact* (i.e., $\widetilde{\boldsymbol{\kappa}}^* = \boldsymbol{\kappa}^*$)
    as $\alpha \to 1$ or as $\alpha \to 0$ or as $\kappa_{\text{tot}} \to 0$.

A2. The appropriate Taylor expansion shows that $\widetilde{\boldsymbol{\kappa}}^* \approx \boldsymbol{\kappa}^*$ if $\alpha^{2\ell} \ll 1$ or if $\kappa_0 \ll 1$ or if
    $\kappa_\ell \ll 1, \forall \ell \geq 1$.

A3. Numerical simulations show that $\widetilde{\boldsymbol{\kappa}}^* \approx \boldsymbol{\kappa}^*$ for moderate values of α, at moderate to high values of κ_{tot} (this is the region where no theoretical justification has been given).

Illustrative examples of the above remarks are given in the sequel.

5.2   Optimal Energy Allocation

The optimal energy vector $\widetilde{\boldsymbol{\kappa}}^*$ is specified by the following:

Theorem. (a) Use the first M data slots $(T_A = M)$ iff

$$\phi_\alpha(M-1) \leq \kappa_{\text{tot}} < \frac{\phi_\alpha(M)}{1 - \delta(M - T + 1)}, \tag{20}$$

where $\delta(x)$ is the Kronecker delta, $1 \leq M \leq T - 1$, and

$$\phi_\alpha(N) = \left[\frac{\alpha^{-N} - 1}{1 - \alpha} - \left(N + \frac{1}{2} \right) + \sqrt{\frac{1}{4} + \frac{(\alpha^{-N} - \alpha)(\alpha^{-N} - 1)}{1 - \alpha^2}} \right].$$

(b) The optimal training energy $(T_A \neq 1)$ is given by:

$$\kappa_0(T_A) = -\Delta(T_A + \kappa_{\text{tot}}) + \sqrt{(\Delta^2 + \Delta)(T_A + \kappa_{\text{tot}})^2 - (\Delta + 1)(T_A + \kappa_{\text{tot}})},$$

where $\Delta = \frac{1}{2} \frac{(1 - \alpha)(1 + \alpha^{T_A})}{\alpha - \alpha^{T_A}}$.

(c) The data energies $(T_A \neq 1)$ are given by, $1 \leq \ell \leq T_A$,

$$\kappa_\ell = \alpha^{\ell - 1} \frac{1 - \alpha}{1 - \alpha^{T_A}} [\kappa_{\text{tot}} - \kappa_0(T_A) + T_A] - 1.$$

(d) If $T_A = 1$, $\kappa_0 = \kappa_1 = \kappa_{\text{tot}}/2$.

Proof. See appendix I.

The channel assignment strategy of equation 20 is illustrated in figure 16 for a system with $\kappa_{\text{tot}} = 50$, $T = 21$, and for several values of α. Consider the curve $\phi_{.8}(M)$. The candidate energy line intersects $\phi_{.8}(M)$ between $M = 10$ and $M = 11$. Therefore, $T_A = 11$ is the optimum number of data slots to activate.
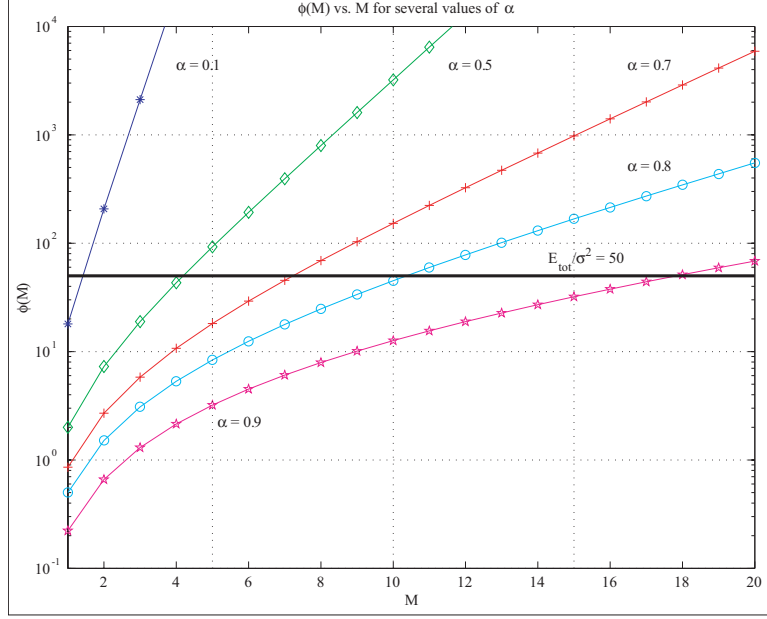
Figure 16. An illustration of the test for determining T_A.

We look at some consequences of the Theorem:

1. T_A is an increasing function of α (see figure 4). This can be verified by noting that

$$\frac{\partial \phi_\alpha(M)}{\partial \alpha} < 0, \quad \text{for } 0 < \alpha < 1.$$

2. As $\kappa_{\text{tot}} \to \infty$, all $T-1$ slots become active and

$$\kappa_0(T-1) = \kappa_{\text{tot}} \left[\frac{\sqrt{h(\alpha)} - 1}{h(\alpha) - 1} \right],$$

where $h(\alpha) = \frac{(1+\alpha)(1-\alpha^{T-1})}{(1-\alpha)(1+\alpha^{T-1})}$.

3. As $\kappa_{\text{tot}} \to 0$, only the first data slot is active and

$$\kappa_0(1) = \frac{\kappa_{\text{tot}}}{2}.$$

5.3 Numerical Simulations

In this section, we show that $\widetilde{\boldsymbol{\kappa}}^* \approx \boldsymbol{\kappa}^*$ using numerical techniques. Define the normalized error metric[‡]

$$\mathbf{e} \triangleq \frac{\|\boldsymbol{\kappa} - \widetilde{\boldsymbol{\kappa}}\|_1}{\kappa_{\text{tot}}}.$$

[‡]Here $\|\mathbf{a}\|_1$ denotes the 1-norm of the vector $\mathbf{a}$.

Figure 17 compares $\widetilde{\boldsymbol{\kappa}}^*$ with $\boldsymbol{\kappa}^*$ for $\kappa_{\mathrm{tot}} = 10$ and $T = 6$, and for several values of α. Remarks A1 and A2 predict that the approximate solution performs well for $\alpha = 0.2$ and $\alpha = 0.98$. This is verified in the figure, both graphically and from the $\mathbf{e}$ metric. We observe that the solution is also close for $\alpha = 0.5$ and $\alpha = 0.7$. Note that in all cases the approximate solution correctly predicts the number of active slots T_A.

In figure 18, we compare $\widetilde{\boldsymbol{\kappa}}^*$ and $\boldsymbol{\kappa}^*$ for $\alpha = 0.85$, $T = 6$, and for different values of κ_{tot}. Remarks A1 and A2 predict accuracy for $\kappa_{\mathrm{tot}} = 0.1$. We see that the normalized error $\mathbf{e}$ remains small for the higher values of κ_{tot} as well. Again, the approximate solution correctly predicts T_A in each case.
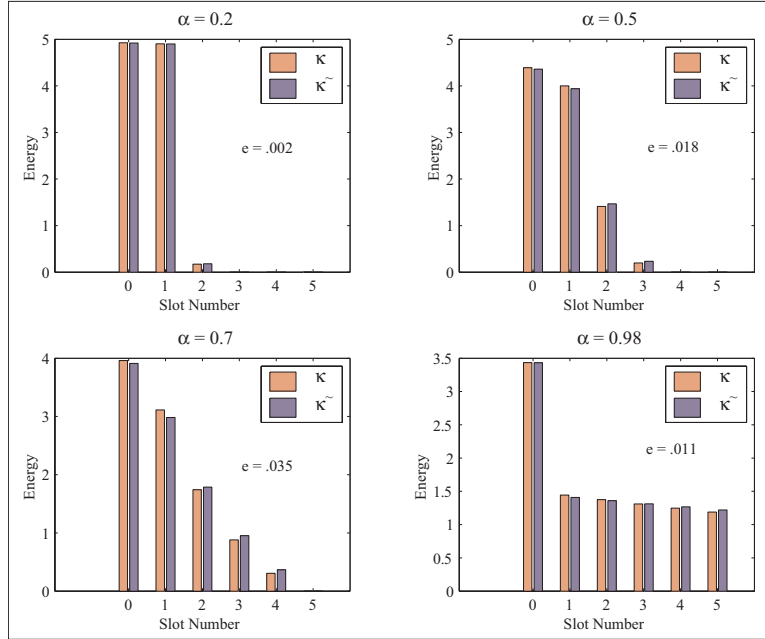
Figure 17. Comparison of $\boldsymbol{\kappa}^*$ with $\widetilde{\boldsymbol{\kappa}}^*$ for $\kappa_{\mathrm{tot}} = 10$ and different values of α. The x-axis indicates each of the T slots. The y-axis shows the energy placed in each of the slots.

5.4 Effect Upon Cutoff Rate

In this section we assess the value of variable data energy allocation over other energy allocation strategies using the cutoff rate metric. In figure 19, we plot: (a) $R_{(o,A)}$, the cutoff rate when all transmission slots use the same energy, so that $\kappa_0 = \kappa_1 = \kappa_{\mathrm{av}}$; (b) $R_{(o,B)}$, the cutoff rate for equal energy data slots, κ_0 is determined from equation 13; (c) $R_{(o,C)}$, the variable data energy cutoff rate using the approximate optimal energy distribution given by the substitution function equation 20, and (d) $R_{(o,D)}$, the variable data energy cutoff rate using the true optimal energy vector determined numerically. The simulation is for $\alpha = 0.98$, and in all cases, the optimal value of the training period T is used.
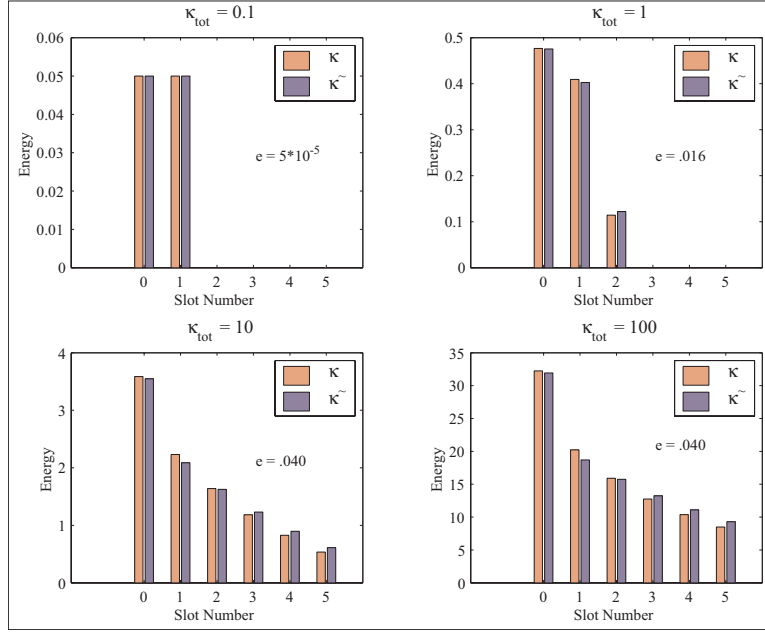
Figure 18. Comparison of $\boldsymbol{\kappa}^*$ with $\widetilde{\boldsymbol{\kappa}}^*$ for $\alpha = 0.85$ and different values of κ_{tot}.
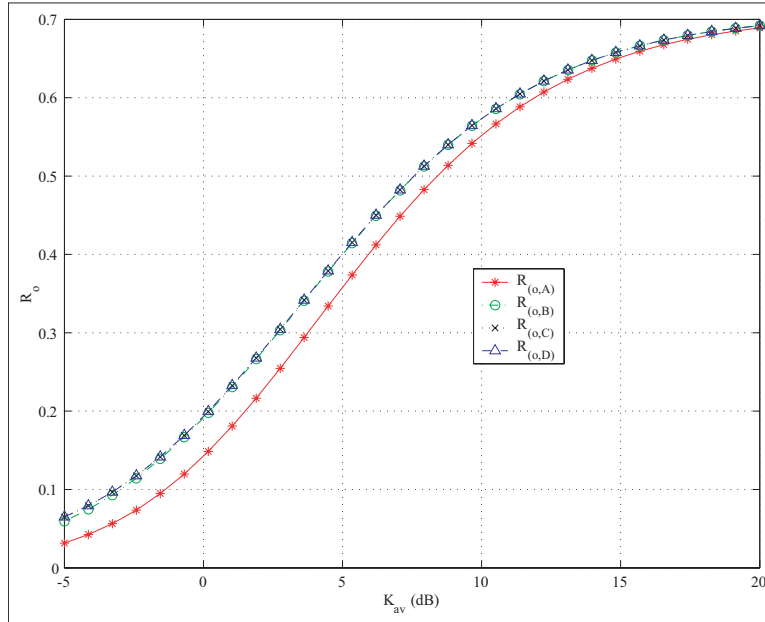
Figure 19. The cutoff rate for four different energy allocation strategies: $R_{(o,A)}$ is the equal-energy cutoff rate, $R_{(o,B)}$ is the two-dimensional cutoff rate, $R_{(o,C)}$ is the variable energy cutoff rate using the substitution function, and $R_{(o,D)}$ is the variable energy cutoff rate using numerical optimization. The Doppler parameter $\alpha = 0.98$, and the optimal value of T is used in all cases.

From figure 19 we make two key observations. First, there is about a 2 dB gain in going from $R_{(o,A)}$ to $R_{(o,B)}$ at low to moderate values of κ_{av}. That is, two dimensional energy optimization results in significant energy savings over an equal energy strategy. This gain increases for larger values of α (slowly fading channels) and diminishes for smaller values of α (rapidly fading channels). As $\kappa_{\mathrm{av}} \to \infty$, this gain diminishes to zero. This is because, in the high energy scenario, even a sloppy allocation of energy will be sufficient for the cutoff rate to saturate. Second, we note that the gain in $R_{(o,C)}$ or $R_{(o,D)}$ over $R_{(o,B)}$ is negligible. This means that *there is practically no benefit in doing variable data energy allocation in place of the fixed data energy allocation.* This result holds true for all tested values of α and κ_{av}.

We also note that these results hold true only if the optimal value of T can be chosen. If the transmitter is unable to choose the optimal value of T, the more sophisticated energy allocation strategies do indeed provide gains as illustrated in figure 20. Here, we let $\alpha = 0.88$ and fix $T = 15$ for all strategies. Note that there is as much as an additional 2 dB gain in going from $R_{(o,B)}$ to $R_{(o,C)}$ (or $R_{(o,D)}$). This is because, in this example, channel predictability is poor. Scheme 'C' (or equivalently 'D') allows the later data slots to be "turned off" since the channel will not be predicted accurately in those slots. The energy saved is used instead in the earlier data slots in which the channel *is* predicted accurately. Schemes 'A' and 'B' do not have this freedom; they are forced to allocate the same energy to all data slots. Also, note that $R_{(o,C)} \approx R_{(o,D)}$. That is, the energy allocation derived from the substitution function results in a cutoff rate that is practically the same as if the exact optimal energy vector had been used.

6. BPSK and OOK Hybrid Modulation

We ask the following question: "Given partial CSI ω_ℓ at the receiver, what is the optimal binary input distribution *in each data slot*?" In light of the discussion in section 3.3, we provide a partial answer by confining our interest to BPSK (optimal for full CSI) and OOK (optimal for no CSI). We will consider the form of OOK for which $p_\ell = 1/2$. The discussion and results in this section are based largely on our results in *[29]*, where proofs had been omitted due to lack of space. Next, we provide the transitional value of the faded SNR $\bar{\kappa}_\ell = g(\omega_\ell)$, above which OOK is optimal, and below which BPSK is optimal:

Design Rule The transitional faded SNR $\bar{\kappa}_\ell$ for the ℓth path is found by equating equation 11 with 12 and solving for κ_1. This yields a third-order polynomial. Retaining the relevant root yields the transitional SNR in the ℓth path to be:

$$\bar{\kappa}_\ell = g(\omega_\ell),$$
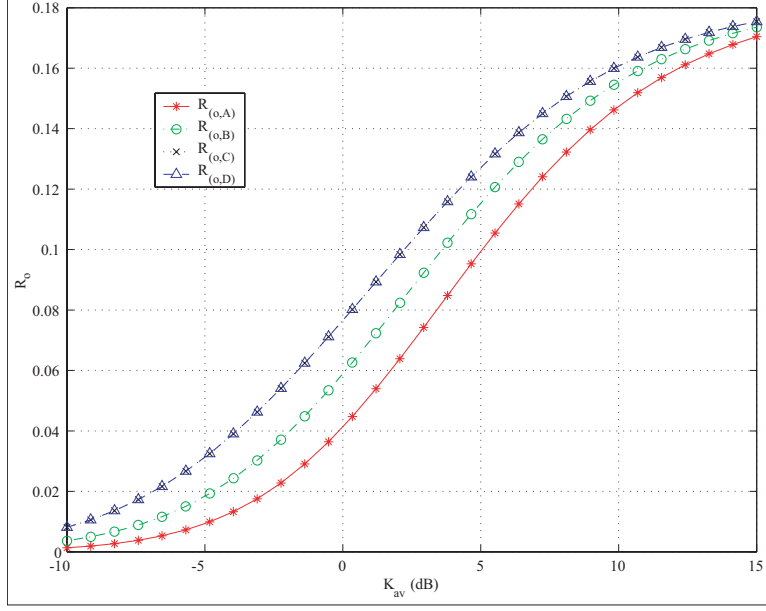
Figure 20. The cutoff rate for four different energy allocation strategies: $R_{(o,A)}$ is the equal-energy cutoff rate, $R_{(o,B)}$ is the two-dimensional cutoff rate, $R_{(o,C)}$ is the variable energy cutoff rate using the substitution function, and $R_{(o,D)}$ is the variable energy cutoff rate using numerical optimization. The Doppler parameter $\alpha = 0.88$, and the value of the training period is fixed, $T = 15$.

where

$$g(\omega_\ell) = \left[\frac{(a+b)^{1/3} + (a-b)^{1/3} - 2\left(4 - 10\omega_\ell + 3\omega_\ell^2\right)}{3(2-\omega_\ell)^2(1-\omega_\ell)} \right],$$

and where

$$a = 81\omega_\ell^6 - 468\omega_\ell^5 + 828\omega_\ell^4 - 640\omega_\ell^3 + 624\omega_\ell^2 - 192\omega_\ell + 64,$$

$$b = 6\sqrt{3}(\omega_\ell - 2)^2\omega_\ell^2\sqrt{61\omega_\ell^4 - 208\omega_\ell^3 + 168\omega_\ell^2 - 64\omega_\ell + 16}. \tag{21}$$

The function $g(\omega_\ell)$ depends on the estimator quality, and is shown in figure 21. At the end points, $\omega_\ell = \{0, 1\}$, our results agree with existing theory:

1. Observe that $g(0) = 0$. Therefore, when no CSI is available, it is always better to use OOK instead of BPSK. This is in agreement with the results of [1], which are for the $\omega_\ell = 0$ case.

2. It can be verified that

$$\lim_{\omega \to 1} \bar{\kappa}_\ell = \infty,$$

which confirms that, when full CSI is available, BPSK is always optimal independent of the faded SNR. This is in agreement with the well known fact that, for AWGN channels, the use of OOK is suboptimal to BPSK for a fixed average symbol energy constraint.

30
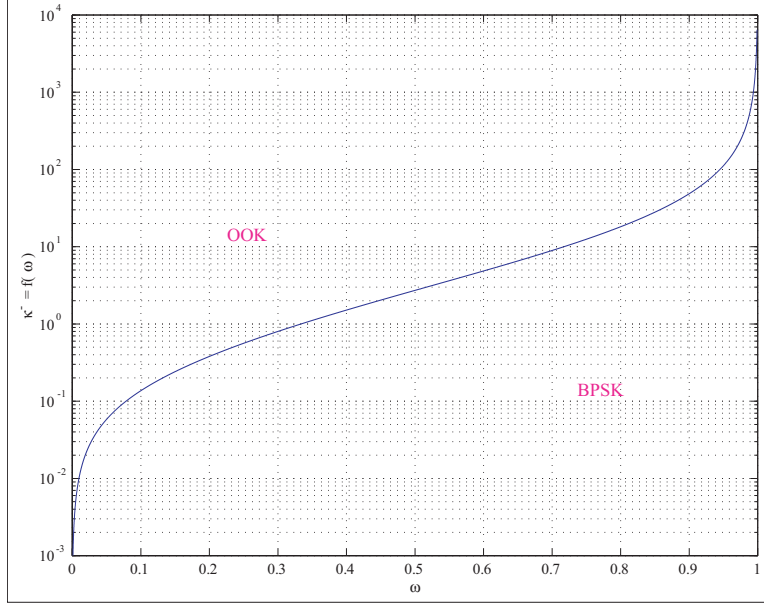
Figure 21. The transitional faded SNR $\bar{\kappa} = f(\omega)$. For larger faded SNR κ OOK is cutoff-rate optimal, and for smaller faded SNR, BPSK is optimal.

3. As expected from figure 21, it can be verified that $g(\omega_\ell)$ is a decreasing function of ω_ℓ. For a fixed κ_ℓ the transmitter should switch from BPSK to OOK as CSI diminishes.

The design rule of equation 21 gives an *analytic basis* for a hybrid modulation scheme in which the transmitter can select between BPSK and OOK based on the faded SNR κ_ℓ and estimator quality ω_ℓ available at the receiver. In [26], the authors used capacity as a metric and considered a similar analysis, where the transmitter was free to choose the optimal binary distribution for each sub-channel (among all possible binary distributions). Because of the intractability of the capacity metric and the input design rule, a numerical analysis was given. Here, we consider a scheme that alternates between OOK and BPSK using only an analytic design rule. The cutoff rate for the BPSK/OOK hybrid modulation scheme is

$$R_o = -\frac{1}{T} \sum_{\mathcal{L}} \log_2 \left\{ \frac{1}{2} + \frac{1}{2} \left[\frac{1 + \kappa_1(1 - \omega_\ell)}{1 + \kappa_1} \right] \right\}$$
$$- \frac{1}{T} \sum_{\mathcal{L}^c} \log_2 \left\{ \frac{1}{2} + \frac{1}{2} \left[\frac{\sqrt{1 + 2\kappa_1(1 - \omega_\ell)}}{1 + \kappa_1 \left(1 - \frac{\omega_\ell}{2}\right)} \right] \right\} \tag{22}$$

where

$$\mathcal{L} = \{\ell : \kappa_{\mathrm{av}} \geq g(\omega_\ell)\} \subseteq \{1, \ldots, T-1\} \tag{23}$$

denotes the set of data slots where BPSK is optimal, and where $\mathcal{L}^c$ denotes the set of data slots where OOK is optimal, with $\mathcal{L} \cup \mathcal{L}^c = \{1, \ldots, T-1\}$. For the $\mathbf{Q_{(1,0)}}$ and $\mathbf{Q_{(\infty,0)}}$ estimators, $g(\omega_\ell)$ is a decreasing function of ℓ. Therefore, sub-channels are initially

31

assigned BPSK modulation, eventually the transitional CSI-level is reached, and the remaining sub-channels are assigned OOK modulation. For the $\mathbf{Q_{(1,1)}}$ estimator, $g(\omega_\ell)$ is largest near the end points. The leading and trailing data slots are assigned BPSK, and the middle slots are assigned OOK, as illustrated in figure 22.
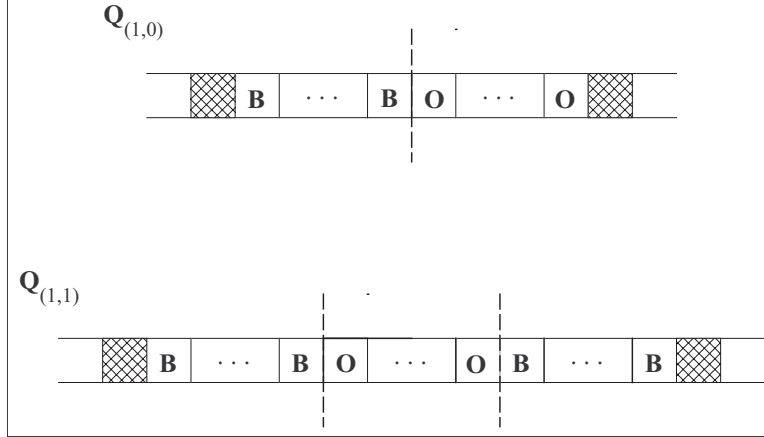
Figure 22. For the causal estimators ($\mathbf{Q_{(1,0)}}$ and $\mathbf{Q_{(\infty,0)}}$), the initial data slots are assigned BPSK, and the latter OOK. For the $\mathbf{Q_{(1,1)}}$ estimator, the beginning and trailing data slots are assigned BPSK, and the intermediary slots are assigned OOK.

We compare our BPSK/OOK adaptive modulation system to:

C1. The BPSK-only system of equation 12. Denote the cutoff rate of this system by $R_{o,\mathrm{BPSK}}$.

C2. The OOK-only system which uses OOK with $p = 1/2$ in each sub-channel. Denote the cutoff rate of this system by $R_{o,\mathrm{OOK}}$; see equation 11.

To simplify the presentation we will consider the equal energy case where $\kappa_0 = \kappa_1 = \kappa_{\mathrm{av}}$. We start by considering the $\mathbf{Q_{(1,0)}}$ estimator in detail. To evaluate equation 22, we first seek to determine $\mathcal{L}$. Evaluating the threshold function yields,

$$\mathcal{L} = \left\{ \ell : \kappa_{\mathrm{av}} \geq g\left(\omega_\ell^{(1,0)}\right) = g\left(\alpha^{2\ell} \frac{\kappa_{\mathrm{av}}}{1 + \kappa_{\mathrm{av}}}\right) \right\}.$$

from which $\mathcal{L}$ can be found explicitly for fixed values of κ_{av} and α (we have added the superscript (1,0) for clarity).

The cutoff rate $R_{o,H}$ for the $\mathbf{Q_{(1,0)}}$ estimator is

$$R_{o,H} = \frac{T-1}{T} - \frac{1}{T}\sum_{\mathcal{L}} \log_2 \left\{ 2 - \alpha^{2\ell} \frac{\kappa_{\text{av}}^2}{(1+\kappa_{\text{av}})^2} \right\}$$

$$- \frac{1}{T}\sum_{\mathcal{L}^c} \log_2 \left\{ 1 + \frac{\sqrt{1 + 2\kappa_{\text{av}}\left(1 - \alpha^{2\ell}\frac{\kappa_{\text{av}}}{1+\kappa_{\text{av}}}\right)}}{1 + \kappa_{\text{av}}\left(1 - \frac{\alpha^{2\ell}}{2}\frac{\kappa_{\text{av}}}{1+\kappa_{\text{av}}}\right)} \right\}.$$

For a fair comparison, each modulation scheme should use the training period T that optimizes its cutoff rate. Indeed, for the OOK based systems, it is possible that sending no training data may be optimal since the cutoff rate is non-zero when $\omega_\ell = 0$. We will consider a system with $\alpha = 0.98$, and use the results of table 2, which show that for a BPSK-only system, the optimal training period at large κ_{tot} is $T = 7$ (the cutoff-rate saturates to 0.71).

In figure 23, we plot the cutoff rate of the BPSK scheme $R_{o,\text{BPSK}}$, the OOK scheme $R_{o,\text{OOK}}$, and the BPSK/OOK adaptive modulation scheme $R_{o,H}$. For small κ_{av}, BPSK is optimal for all sub-channels (from equation 21), and so the cutoff-rate $R_{o,H}$ is equal to $R_{o,\text{BPSK}}$. For the intermediate values of κ_{av}, BPSK is optimal for the initial sub-channels, while OOK is optimal for the latter ones. In this region, $R_{o,H}$ is larger than $R_{o,\text{BPSK}}$. To find the κ_{av} above which $R_{o,H}$ becomes larger than $R_{o,\text{BPSK}}$ we solve for κ_{av} in the following equation

$$\kappa_{\text{av}} = g\left(\omega_{T-1}^{(1,0)}\right) = g\left(\alpha^{2(T-1)}\frac{\kappa_{\text{av}}}{1+\kappa_{\text{av}}}\right), \tag{24}$$

which, with $\alpha = 0.98$ and $T = 7$, indicates that the OOK/BPSK hybrid scheme outperforms the BPSK-only scheme starting at $\kappa_{\text{av}} \approx 9.8$ dB. This is confirmed in the figure.

For large κ_{av}, $R_{o,H}$ is equal to $R_{o,\text{OOK}}$ since OOK outperforms BPSK in all sub-channels. To find the κ_{av} above which the OOK only scheme performs as well as the BPSK/OOK adaptive scheme we solve for κ_{av} in

$$\kappa_{\text{av}} = g\left(\omega_1^{(1,0)}\right) = g\left(\alpha^2\frac{\kappa_{\text{av}}}{1+\kappa_{\text{av}}}\right), \tag{25}$$

which yields the intersection point as $\kappa_{\text{av}} \approx 20.73$ dB.

At any value of the faded SNR κ_{av}, the BPSK/OOK scheme performs *at least* as well as the best of the BPSK or OOK only approaches, and for some intermediate range of κ_{av}, the adaptive scheme performs better than the best of either the OOK only or BPSK only schemes. Note again that substantial gains are obtained by using OOK in the latter slots.

Figure 23. A comparison of the cutoff rate for several modulation techniques for the $\mathbf{Q}_{(\mathbf{1,0})}$ estimator with $\alpha = 0.98$ and $T = 7$. $R_{o,\mathrm{BPSK}}$ denotes BPSK only, $R_{o,\mathrm{OOK}}$ denotes the use of OOK with $p_\ell = 1/2$, and $R_{o,H}$ denotes the BPSK/OOK adaptive modulation scheme.

Next, we repeat the preceding analysis for the $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimator. The hybrid cutoff rate $R_{o,H}$ is again given by equation 22 and $\mathcal{L}$ by equation 23, where the estimator quality is now given by $\omega_\ell^{(\infty,0)}$ (see equation 6). To find the two intersection points, we must solve $\kappa_{\mathrm{av}} = g\left(\omega_{T-1}^{(\infty,0)}\right)$ and $\kappa_{\mathrm{av}} = g\left(\omega_1^{(\infty,0)}\right)$ in a fashion analogous to equations 24 and 25. This yields the intersection points $\kappa_{\mathrm{av}} \approx 10.4$ dB and $\kappa_{\mathrm{av}} \approx 20.8$ dB (because $\omega_\ell^{(\infty,0)} \geq \omega_\ell^{(1,0)}$, the intersection points must be to the right of their values in the previous case). This is confirmed in figure 24, where we plot $R_{o,H}$ for the $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimator for $\alpha = 0.98$ and $T = 7$ (the choice T is again derived from table 2).

Next, we consider the $\mathbf{Q}_{(\mathbf{1,1})}$ estimator in figure 25. The hybrid cutoff rate $R_{o,H}$ is again given by equation 22 and $\mathcal{L}$ by equation 23, where the estimator quality is now given by $\omega_\ell^{(1,1)}$. We take $\alpha = 0.98$ and set $T = 11$ based on table 2. For this *non-causal* estimator, the left intersection point is now found by solving for κ_{av} in the equation $\kappa_{\mathrm{av}} = g\left(\omega_{\lceil \frac{T-1}{2}\rceil}^{(1,1)}\right)$; this yields the value of κ_{av} above which the central data slot will be assigned OOK. To find the right intersection point, we solve $\kappa_{\mathrm{av}} = g\left(\omega_{T-1}^{(1,1)}\right)$; this yields the κ_{av} above which the last (and also the first) data slot is assigned OOK.
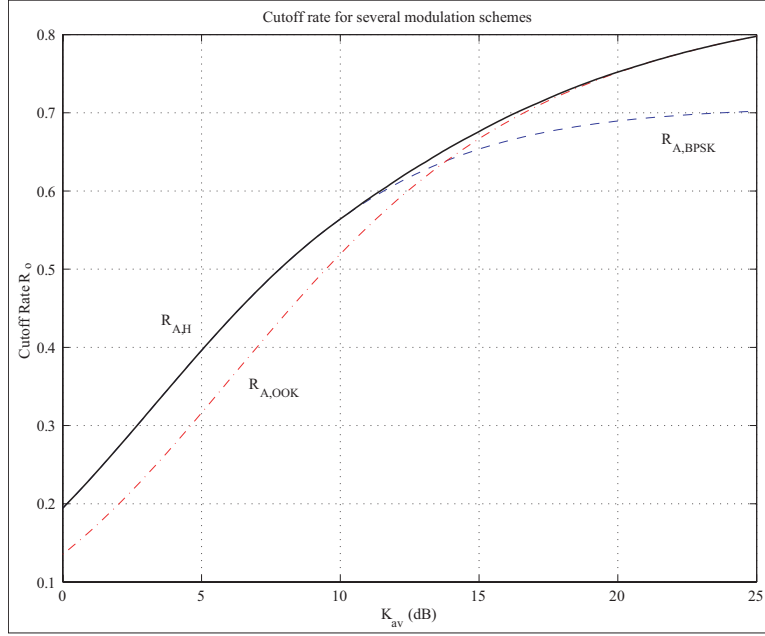
Figure 24. A comparison of the cutoff rate for several modulation techniques for the $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimator with $\alpha = 0.98$ and $T = 7$. $R_{o,\text{BPSK}}$ denotes BPSK only, $R_{o,\text{OOK}}$ denotes the use of OOK with $p_\ell = 1/2$, and $R_{o,H}$ denotes the BPSK/OOK adaptive modulation scheme.
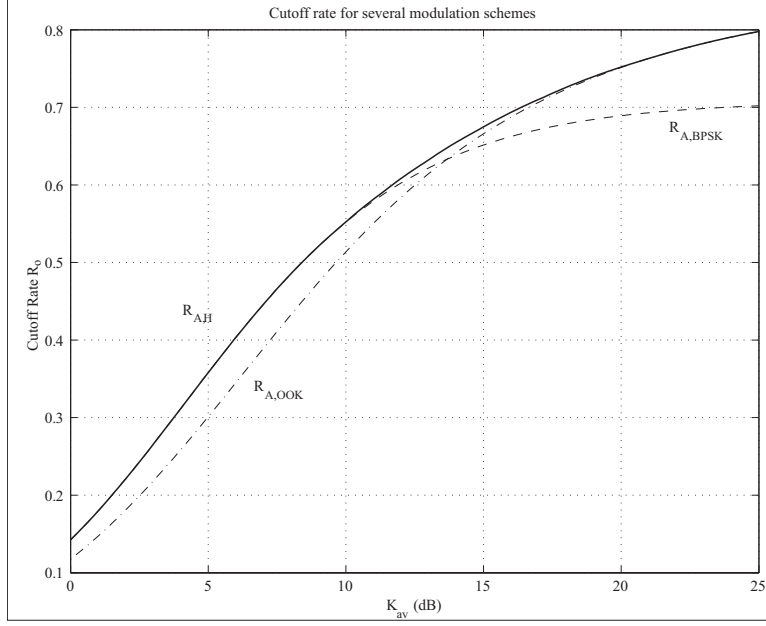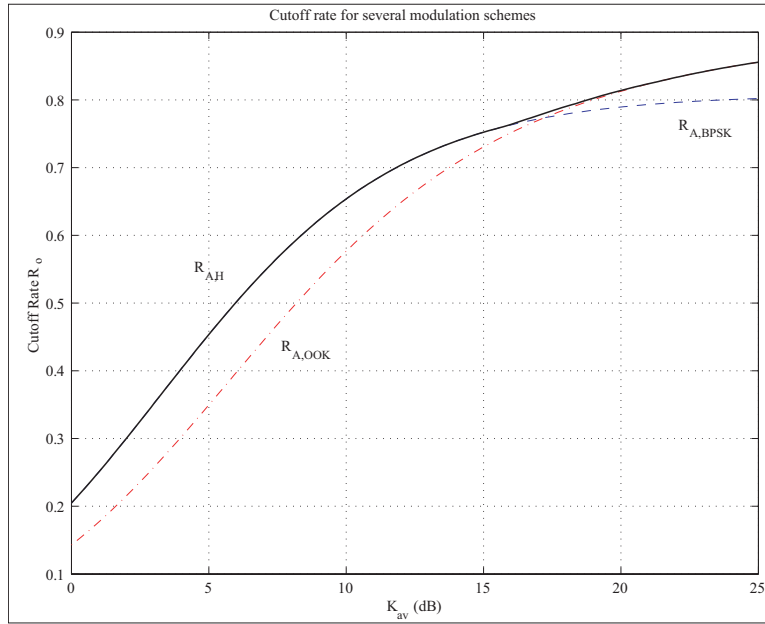

Figure 25. A comparison of the cutoff rate for several modulation techniques for the $\mathbf{Q}_{(\mathbf{1},\mathbf{1})}$ estimator with $\alpha = 0.98$ and $T = 11$. $R_{o,\text{BPSK}}$ denotes BPSK only, $R_{o,\text{OOK}}$ denotes the use of OOK with $p_\ell = 1/2$, and $R_{o,H}$ denotes the BPSK/OOK adaptive modulation scheme.

For each of the three estimators considered, we see that the hybrid scheme captures the optimality of BPSK at low SNR (resulting in a energy savings of up to $2 \sim 3$ dB versus the pure OOK scheme), and the optimality of OOK at high SNR (allowing the cutoff rate to saturate to its maximum possible value of $T/(T-1)$ as $\kappa_{\mathrm{av}} \to \infty$). In addition, there is a region, located between the two intersection points as discussed above, for which the hybrid scheme outperforms the best of the pure OOK and pure BPSK schemes at each value of κ_{av}.

7. Optimal Training for the Jakes Model

In this section, we will consider the cutoff rate for the Jakes channel correlation model [18]. The Jakes model is considered to be an excellent description of the channel correlation of real-world communication channels. However, the Jakes model is often not amenable to analysis, leading to use of autoregressive models (e.g., the Gauss Markov AR(1) model) in its place. In this section, we consider the effect on cutoff rate R_o if the Jakes model describes the channel correlation. We also assess the value of the optimal energy and training period rules designed for the AR(1) model when they are instead applied to the Jakes model. In this section, we will consider BPSK signaling.

The observation under the Jakes model is the same as before, from equation 1,

$$y'_k = \sqrt{E_k} h'_k s_k + n'_k,$$

where h'_k is again a zero-mean complex Gaussian random process with variance σ_h^2, but now has the channel correlation $R_J(\tau) = J_o(2\pi f_D T_D |\tau|)$ where $J_o(.)$ is the zero-th order Bessel function of the first kind, f_D is the maximum Doppler frequency and T_D is the symbol duration. After interleaving, estimation, and de-interleaving, the observation equation becomes (rewriting equation 8)

$$y_k = \sqrt{E_{\lceil k/T \rceil}} \, \widehat{h}_k s_k + \sqrt{E_{\lceil k/T \rceil}} \, \widetilde{h}_k s_k + n_k,$$

and the corresponding cutoff rate for the $\mathbf{Q}_{(\mathbf{x},\mathbf{y})}$ estimator under the Jakes model is given by

$$R_{o,J}^{(x,y)} = -\frac{1}{T} \sum_{\ell=1}^{T-1} \log_2 \left\{ \frac{1}{2} + \frac{1}{2} \left[\frac{1 + \kappa_1(1 - \omega_{\ell,J}^{(x,y)})}{1 + \kappa_1} \right] \right\},$$

where $\omega_{\ell,J}^{(x,y)}$ denotes the estimator quality for the $\mathbf{Q}_{(\mathbf{x},\mathbf{y})}$ estimator, under the Jakes model. In appendix A, the estimator quality equations are found to be (letting $R_h(\tau) = R_J(\tau)$)

$$\omega_{\ell,J}^{(1,0)} = J_o^2(2\pi f_D T_D \ell) \frac{\kappa_0}{1 + \kappa_0},$$

$$\omega_{\ell,J}^{(1,1)} = (\Gamma_{(\ell)}^2 + \Gamma_{(T-\ell)}^2)(\kappa_0^2 + \kappa_0) + 2\kappa_0^2 \Gamma_{(\ell)} \Gamma_{(T-\ell)} J_o(2\pi f_D T_D T),$$

36

where
$$\Gamma_{(k)} \triangleq \frac{J_o\left[2\pi f_D T_D k\right](\kappa_0+1) - J_o\left[2\pi f_D T_D T\right] J_o\left[2\pi f_D T_D(T-k)\right]\kappa_0}{(\kappa_0+1)^2 - \kappa_0^2 J_o^2\left[2\pi f_D T_D T\right]}.$$

Next, we will test our design paradigms for energy allocation and the training period that were derived from the AR(1) model, on the Jakes model. Do our designs obtain near-optimality in the Jakes channel correlation model? If they do, then the results are useful, as the Jakes model is taken to be an excellent model of real-world wireless channels. Next, we will compare the value of the Jakes cutoff rate with the AR(1) cutoff rate. It is desirable that they be in close agreement, as this will validate the insights gained from studying the cutoff rate curves based on the AR(1) model in the previous analysis.

In comparing the Jakes and AR(1) models, an important issue arises: what value of α (used to measure Doppler in the AR(1) model) corresponds to a fixed value of $f_D T_D$ (used to measure Doppler in the Jakes model)? One plausible way to compare models is to use a weighted mean square error distortion metric so that

$$\widehat{\alpha}(T_D f_D, M) = \arg\min_{\alpha} \sum_{\ell=1}^{M} v_\ell \left|R_J(\ell) - R_A(\ell)\right|^2 \tag{26}$$

where $R_A(\tau) = \alpha^{|\tau|}$, $R_J(\tau) = J_o(2\pi f_D T_D \tau)$, M is the number of lags over which we wish to match the two correlation functions, and $\{v_\ell\}_{\ell=1}^{M}, v_\ell \geq 0$ are the weights. This weighting is certainly necessary. For example, for causal estimators, earlier "lags" contribute more to the cutoff rate than later lags. However, there is a problem with this approach. Note that $\alpha(T_D f_D, M)$ changes as the number of lags of interest (M) changes. More importantly, how do we determine the value of the weights v_ℓ in equation 26?

Here, we will not attempt to design a universal mapping rule. Instead, we will require that each value of the pair $(f_D T_D, T)$ should be mapped to one value of α, i.e., for all κ_{av}. We will be omniscient in how this value of α is chosen. We will always pick the "best-fit" α, as will be described in the sequel.

7.1 Energy Allocation

Here, we compare the optimal energy allocation $(\kappa_{0,J}^*, \kappa_{1,J}^*)$ when the Jakes model is used to that when the AR(1) model is used $(\kappa_{0,A}^*, \kappa_{1,A}^*)$. As before, we impose a total energy constraint $\kappa_0 + (T-1)\kappa_1 \leq \kappa_{\text{av}} T$ and consider the value of the training period to be fixed (the optimal training period is considered in the next section).

For the $\mathbf{Q}_{(1,0)}$ estimator, it is easy to verify that the optimal energy allocation is the same for both the Jakes and AR(1) models. This is because, for this estimator, the optimal energy allocation does not depend on the channel correlation function at all. As a special

case, we note that this result also holds for the purely bandlimited fading process considered in [31]. The optimal data energy is given by (see equation 13)

$$\kappa_{1,J}^* = \kappa_{0,J}^* = \Gamma - \sqrt{\Gamma^2 - \frac{\kappa_{\text{tot}}}{T-1}\Gamma}, \quad \Gamma = \frac{\kappa_{\text{av}}T + 1}{T - 2},$$

for $T > 2$. For $T = 2$, $\kappa_{1,J}^* = \kappa_{1,A}^* = \kappa_{\text{tot}}/2$. For this simple estimator, the AR(1) model correctly predicts the optimal energy allocation for the Jakes model.

Next, we consider the $\mathbf{Q}_{(1,1)}$ estimator. In figure 26, we plot the cutoff rate for both the Jakes model and for the AR(1) model for two different values of the Doppler spread. We choose a carrier frequency of 900 MHz and maximum Doppler spreads $f_D = 25$ Hz (with $T = 15$) and $f_D = 100$ Hz (with $T = 5$) (this corresponds to mobile speeds of 30 km/hr and 120 km/hr, respectively [32, pp.141-143]). The symbol period T_D is 1 msec. At each value of κ_{tot}, the cutoff rate is optimized over the energy allocation (κ_0, κ_1) separately for each model. For the AR(1) model, the best-fit α is found by finding the single value of α that minimizes the average difference in the cutoff rates $|R_{o,J} - R_{o,A}|$ over all κ_{tot}. We find that $f_D T_D = 0.1$ corresponds to $\alpha = 0.88$ and that $f_D T_D = 0.025$ corresponds to $\alpha = 0.99$. From the figure, we see that, when the appropriate value of α is chosen, the cutoff rate of the Jakes model closely matches that of the AR(1) model over all κ_{tot}.

We note that the best-fit value of α is chosen using the $|R_{o,J} - R_{o,A}|$ criteria, and thus, the difference in the associated training energies $|\kappa_{0,J} - \kappa_{0,A}^*|$ may be large, as is evident from figure 27. Next, we ask the following question: If we allocate training energy to the Jakes model based on the optimal training energy for the AR(1) model (i.e., we let $\kappa_{0,J} = \kappa_{0,A}^*$), what is the effect upon cutoff rate? In figure 28, we plot the cutoff rate for the Jakes model when $\kappa_{0,J} = \kappa_{0,J}^*$ (denote this cutoff rate by $R_{o,J}^*$) and when $\kappa_{0,J} = \kappa_{0,A}^*$ (denote this cutoff rate by $R_{o,J}$). The system parameters are the same as in the previous case. We see that there is virtually no loss in the cutoff rate when the energy allocation based on the AR(1) model is used to dictate the energy allocation for the Jakes model, provided that an appropriate value of α is chosen.

7.2   Training Period

In this section we consider the optimal training period T^* under the Jakes model. We will assume that $f_D T_D \ll 1$, so that we are interested only in the first decreasing "half-lobe" of the Jakes function. In appendix G, it is shown that for the $\mathbf{Q}_{(1,0)}$ estimator, a lower bound on the optimal training period in the high SNR ($\kappa_{\text{tot}} \to \infty$) scenario, under the Jakes model, is again given by

$$T_{B,(1,0)} = \arg\min_T \prod_{\ell=1}^{T-1} \left[1 - \frac{J_o^2(2\pi f_D T_D \ell)}{2} \right]^{1/T}. \tag{27}$$

38
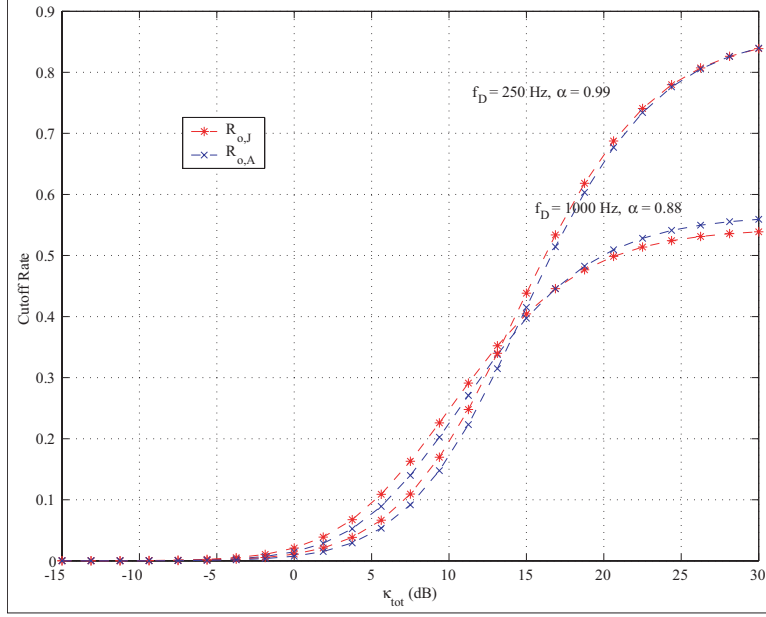
Figure 26. The cutoff rate for the Jakes model $R_{o,J}$ and AR(1) model $R_{o,A}$ for two sets of parameters: $f_D = 25$ Hz, $\alpha = 0.99$, $T = 15$ and $f_D = 100$ Hz, $\alpha = 0.88$, $T = 5$. The symbol period is $T_D = 1$ msec and the carrier frequency $f_c = 900$ MHz.
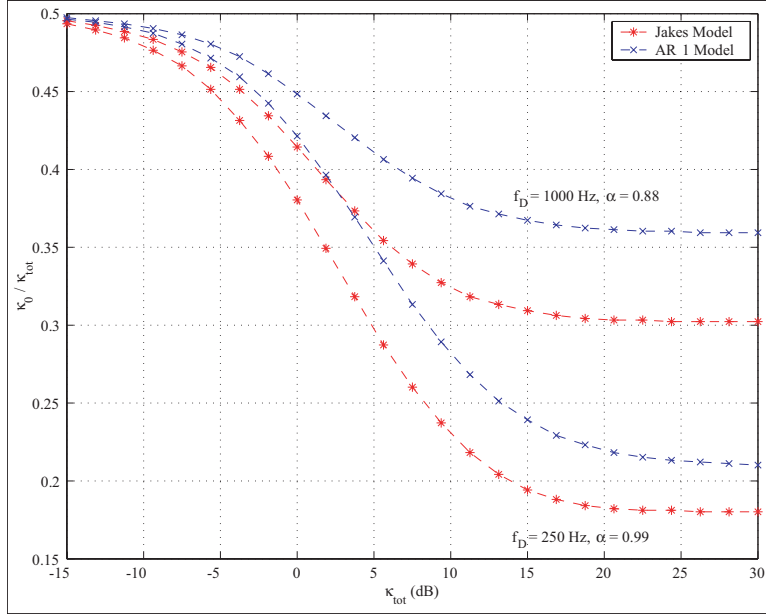
Figure 27. The training energy ratio $\kappa_{J,0}^*/\kappa_{\text{tot}}$ for the Jakes model and AR(1) model for two sets of parameters: $f_D = 25$ Hz, $\alpha = 0.99$, $T = 15$ and $f_D = 100$ Hz, $\alpha = 0.88$, $T = 5$. The symbol period is $T_D = 1$ msec and the carrier frequency $f_c = 900$ MHz.

Also in the appendix, it is shown that this lower bound is exact at high SNR, and valid for any channel correlation function that decreases monotonically in the range of interest.

39
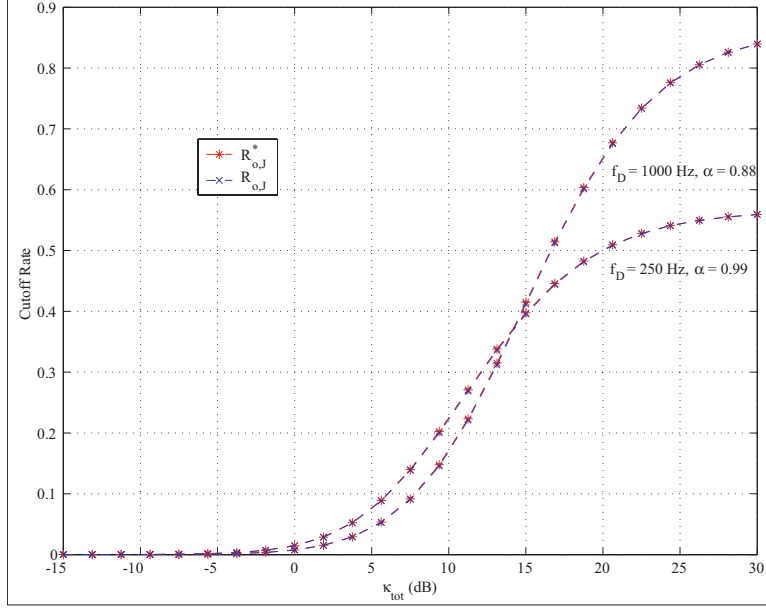
Figure 28. The cutoff rate for the Jakes model for two sets of parameters: $f_D = 25$ Hz, $\alpha = 0.99$, $T = 15$ and $f_D = 100$ Hz, $\alpha = 0.88$, $T = 5$. $R^*_{o,J}$ denotes the cutoff rate using the true optimal energy allocation. $R_{o,J}$ denotes the cutoff rate when the optimal energy allocation *based on the AR(1) model* is used instead.

Even without the assumption $f_D T_D \ll 1$, numerical evidence indicates that equation 27 is valid. The lower bound is illustrated in table 3. Again, we choose $T_D = 1$ msec and a carrier frequency of 900 MHz. The values of $f_D T_D$ shown in the table are 0.1, 0.05, 0.025, and 0.01 and correspond to mobile speeds of 120, 60, 30, and 12 km/hr, respectively. If we approximate $f_D T_D = 0.1$ and $f_D T_D = 0.01$ in the Jakes model as being equivalent to $\alpha = 0.80$ and $\alpha = 0.99$ in the AR(1) model (based on our analysis in the previous subsection), we see that both the AR(1) model and Jakes model result in very similar training period designs. The value of the lower bound is nearly the same for the two models. Additionally, we see that in both models, the bound is tight even at low to moderate SNR, with this tightness increasing as the Doppler spread increases. Therefore, we conclude that for the $\mathbf{Q}_{(\mathbf{1},\mathbf{0})}$ estimator, the training period results derived from the AR(1) model are also applicable when the Jakes model describes the channel correlation.

Based on the preceding discussion, we conclude that an analysis of the AR(1) model results in useful insights on how to allocate energy, choose the training period, and determine a range of reliable transmission rates. Furthermore, these results are useful even if the "true" wireless channel (i.e., the channel encountered in practice) is described well by the Jakes model.

Table 3. Comparison of the optimal training period $T^*_{(1,0)}$ to the lower bound $T_{B,(1,0)}$ for several different values of $f_D T_D$ and κ_{av} under the Jakes model.

	$T^*_{(1,0)}$	$T_{B,(1,0)}$
$f_D T_D = 0.1$		
$\kappa_{\mathrm{av}} = 1$	3	**2**
$\kappa_{\mathrm{av}} = 10$	2	**2**
$\kappa_{\mathrm{av}} = 100$	2	**2**
$f_D T_D = 0.05$		
$\kappa_{\mathrm{av}} = 1$	4	**3**
$\kappa_{\mathrm{av}} = 10$	4	**3**
$\kappa_{\mathrm{av}} = 100$	3	**3**
$f_D T_D = 0.025$		
$\kappa_{\mathrm{av}} = 1$	7	**5**
$\kappa_{\mathrm{av}} = 10$	5	**5**
$\kappa_{\mathrm{av}} = 100$	5	**5**
$f_D T_D = 0.01$		
$\kappa_{\mathrm{av}} = 1$	14	**9**
$\kappa_{\mathrm{av}} = 10$	10	**9**
$\kappa_{\mathrm{av}} = 100$	9	**9**

So far we have considered two particular channel correlation models (the AR(1) and Jakes) and three MMSE estimators ($\mathbf{Q_{(1,0)}}$, $\mathbf{Q_{(\infty,0)}}$, and $\mathbf{Q_{(1,1)}}$). In the next section, we generalize the framework to an arbitrary channel correlation function $R_h(\tau) = \frac{1}{\sigma_h^2} \mathcal{E}\left[h'_k h'^{*}_{k+\tau} \right]$ and for an entire class of MMSE estimators.

---

8. Generalized Channel and Estimation Model

In this section, we will generalize the channel correlation model and the channel estimation model to show how our previous results follow as special cases of a more general system model. For the sake of completeness, we restate the channel observation, state, and estimation expressions in general terms. We will, however, consider only binary input strategies and consider the case where all data transmissions are assigned equal energy.

The channel observation is given by

$$y'_k = \sqrt{E_k} h'_k s_k + n'_k,$$

where the coded input s_k is selected from a binary signal set $\mathcal{S} = \{A, B\}$ (i.e., $s_k \in \mathcal{S}$) and subject to a unit average-energy constraint: $pA^2 + (1-p)B^2 = 1$, where p is the probability

of transmitting A. The sequence $n'_k \sim \mathcal{CN}(0, \sigma_N^2)$ describes AWGN, and the transmission energy used at time k is $E_k|s_k|^2$. The Gaussian channel $h'_k \sim \mathcal{CN}(0, \sigma_h^2)$ has a general correlation function $R_h(\tau) \triangleq \frac{1}{\sigma_h^2} \mathcal{E}[h'_k h'^*_{k+\tau}]$ and describes time-correlated Rayleigh fading. Codewords of length $N' = N(T-1), N \in \mathcal{Z}$ are used.

Training is sent periodically once every T transmissions, i.e., during the time slots $k = mT$. At each time $mT + \ell$, an MMSE estimate of the channel $\widehat{h}'_{mT+\ell}$ is made at the receiver using some subset $\mathcal{N}$ of past (and possibly future) training symbol observations, so that

$$\widehat{h}'_{mT+\ell} = \mathcal{E}[h'_{mT+\ell} | \{y'_{nT}\}, \ n \in \mathcal{N} \subset \mathcal{Z}]. \tag{28}$$

The use of an MMSE estimator implies that $\widehat{h}'_{mT+\ell} \sim \mathcal{CN}(0, \widehat{\sigma}_\ell^2)$, where $\widehat{\sigma}_\ell^2$ is the estimator variance. From orthogonality, that $\widetilde{h}'_{mT+\ell} \sim \mathcal{CN}(0, \sigma_h^2 - \widehat{\sigma}_\ell^2)$. That is, $\widehat{h}'_{mT+\ell}$ and $\widetilde{h}'_{mT+\ell}$ are independent. To characterize the performance of a particular estimator, we will define the *estimator quality* ω_ℓ as

$$\omega_\ell \triangleq \widehat{\sigma}_\ell^2 / \sigma_h^2. \tag{29}$$

Note that orthogonality implies that $0 \leq \omega_\ell \leq 1$.

Denoting the estimation error at time k by $\widetilde{h}'_k$, the system equation can be written as

$$\begin{aligned} y'_k &= \sqrt{E_{\lceil k/T \rceil}} \ h'_k s_k + n'_k \\ &= \sqrt{E_{\lceil k/T \rceil}} \ \widehat{h}'_k s_k + \sqrt{E_{\lceil k/T \rceil}} \ \widetilde{h}'_k s_k + n'_k, \end{aligned}$$

where we have assumed that the energy allocation for all data transmissions E_1 is constant due to practical constraints (e.g., peak-to-average-power ratio specifications and transmitter complexity). The input s_k is now selected from a complex signal set $\mathcal{S}_{[k]} = \{A_{[k]}, B_{[k]}\}$ and subject to a unit average-energy constraint: $p_{[k]} A_{[k]}^2 + (1 - p_{[k]}) B_{[k]}^2 = 1 \ \forall k \neq mT$. In the training slots, we assume, without loss of estimator performance, that $\mathcal{S}_0 = \{+1\}$.

We assume that perfect interleaving is performed at the transmitter, and that channel estimation is done before de-interleaving at the receiver. The observation equation is now

$$y_k = \sqrt{E_k} h_k s_k + n_k,$$

where h_k is an i.i.d. sequence representing the interleaved channel, and where $n_k \sim \mathcal{CN}(0, \sigma_N^2)$ is another AWGN sequence. Writing this in terms of the channel estimate and error

$$\begin{aligned} y_k &= \sqrt{E_{\lceil k/T \rceil}} \ h_k s_k + n_k \\ &= \sqrt{E_{\lceil k/T \rceil}} \ \widehat{h}_k s_k + \sqrt{E_{\lceil k/T \rceil}} \ \widetilde{h}_k s_k + n_k. \end{aligned} \tag{30}$$

Interleaving implies that $\widehat{h}_k$ and $\widetilde{h}_k$ are independent sequences in k, and also with respect to each other. However, interleaving does not change the marginal statistics of the channel

estimate and estimation error: $\widehat{h}_{mT+\ell} \sim \mathcal{CN}\left(0, \widehat{\sigma}_\ell^2\right)$ and $\widetilde{h}_{mT+\ell} \sim \mathcal{CN}\left(0, \sigma_h^2 - \widehat{\sigma}_\ell^2\right)$. Furthermore, $\widehat{h}_{mT+\ell}$ and $\widetilde{h}_{mT+\ell}$ are independent. The estimator quality is defined by equation 29 as before. Finally, we assume that codewords are decoded using the ML-detector which treats $s_{mT+\ell}$ as the channel input and the pair $(y_{mT+\ell}, \widehat{h}_{mT+\ell})$ as the channel output.

The cutoff rate for generalized binary transmission, given the front-end above, is given by (see appendix D with $\kappa_\ell = \kappa_1$)

$$R_o = -\frac{1}{T}\sum_{\ell=1}^{T-1}\log_2 \min_{p_\ell A_\ell^2 + (1-p_\ell)B_\ell^2 = 1}$$

$$\left[1 + 2p_\ell\left(1-p_\ell\right)\left\{\frac{\sqrt{1+\kappa_1\left(1-\omega_\ell\right)|A_\ell|^2}\sqrt{1+\kappa_1\left(1-\omega_\ell\right)|B_\ell|^2}}{1+\frac{1}{2}\kappa_1\left(1-\omega_\ell\right)\left(|A_\ell|^2+|B_\ell|^2\right)+\frac{1}{4}\kappa_1\,\omega_\ell\,|A_\ell - B_\ell|^2}-1\right\}\right],$$

where $\kappa_1 \triangleq \sigma_h^2 E_1/\sigma_N^2$ was previously defined as the faded SNR during data transmissions. Additionally, we define $\kappa_0 \triangleq \sigma_h^2 E_0/\sigma_N^2$ as the faded training energy.

Next, we compute the estimator quality for three particular estimators (derivations are given in appendix A) Consider the following estimators for $1 \le \ell \le T-1$:

G1. The $\mathbf{Q}_{(1,0)}$ estimator for which $\mathcal{N} = \{m\}$. The channel estimate and the estimator quality $\omega_\ell^{(1,0)}$ are given by

$$\widehat{h}'_{mT+\ell} = \frac{\sqrt{E_0}\sigma_h^2}{E_0\sigma_h^2 + \sigma_N^2}R_h(\ell)y'_{mT},$$

$$\omega_\ell^{(1,0)} = R_h^2(\ell)\frac{\kappa_0}{1+\kappa_0}.$$

G2. The $\mathbf{Q}_{(\infty,0)}$ estimator for which $\mathcal{N} = \{m, m-1, m-2, \ldots\}$. The channel estimate and the estimator quality $\omega_\ell^{(1,0)}$ are given by

$$\widehat{h}'_{mT+\ell} = \frac{1}{\sqrt{E_o}}\sum_{v=0}^{\infty}y'_{vT+\ell}\gamma_{v,\ell} \text{ and}$$

$$\omega_\ell^{(\infty,0)} = \frac{1}{\sigma_h^2}\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}R_h(nT+\ell)R_h(mT+\ell)z(|n-m|)$$

where $z(\tau) \triangleq \mathcal{Z}^{-1}\left\{\frac{1}{\mathcal{Z}[R_h(T\tau)]+\frac{1}{\kappa_0}}\right\}$ and $\gamma_{v,\ell} \triangleq \sum_{n=0}^{\infty}R_h(nT+\ell)z(|n-v|)$.

G3. The $\mathbf{Q}_{(1,1)}$ for which $\mathcal{N} = \{m, m+1\}$. The channel estimate and the estimator quality $\omega_\ell^{(1,0)}$ are given by

$$\widehat{h}'_{mT+\ell} = \frac{\sigma_h^2}{\sigma_N^2} \sqrt{E_0} \left[\Gamma_{(\ell)} \, y'_{mT} + \Gamma_{(T-\ell)} \, y'_{(m+1)T} \right],$$

$$\omega_\ell^{(1,1)} = (\Gamma_{(\ell)}^2 + \Gamma_{(T-\ell)}^2)(\kappa_0^2 + \kappa_0) + 2\kappa_0^2 \Gamma_{(\ell)} \Gamma_{(T-\ell)} R_h(T),$$

where

$$\Gamma_{(k)} \triangleq \frac{R_h(k)(\kappa_0 + 1) - R_h(T) R_h(T-k)\kappa_0}{(\kappa_0 + 1)^2 - \kappa_0^2 \, R_h^2(T)}.$$

In this section, we have shown how to formulate the cutoff rate for a general channel correlation matrix $R_h(\tau)$ and for a general MMSE estimator. By letting $R_h(\tau) = R_A(\tau) = \alpha^{|\tau|}$ we are able to derive the results for the AR(1) model. Similarly, by letting $R_h(\tau) = R_J(\tau) = J_o(2\pi f_d T_D |\tau|)$ we are able to derive the results for the Jakes model. Additionally, we see that it is easy to consider an entire class of MMSE estimators for which our cutoff rate metric holds, each defined by the pilot set $\mathcal{N}$, or equivalently, it's estimator quality ω_ℓ.

9. Discussion and Summary

In this report we have considered the optimal allocation of resources (training energy and training frequency) for correlated fading channels when partial CSI is available at the receiver through periodic training. We have used the channel cutoff rate as our optimization metric. Mainly, we have assumed that the transmitter has perfect knowledge of the channel Doppler spread, but have also treated the case where the transmitter's knowledge is incorrect.

First, we reviewed the Gauss-Markov correlated fading channel and discussed a periodic training-based channel estimation scheme which provides partial CSI to the receiver by taking an MMSE estimate of the channel based on some subset of the received pilots. Three different MMSE estimators (i.e., three different pilot subsets) were considered, and the characteristic *estimator quality* of each estimator was given.

Next, we derived the cutoff rate R_o for our training-based front end. Although our emphasis is on the Gauss-Markov channel, this cutoff rate metric (which is parameterized by the estimator quality, SNR, and training period) holds for *any* channel correlation function. We focused on binary signaling and noted that OOK is optimal when no CSI is available, whereas BPSK is optimal when full CSI is available. To study the intermediate region of partial CSI, we derived a design rule which gives, as a function of the CSI and

SNR available at the receiver, the optimal input distribution (OOK or BPSK). We have shown that adaptively switching from BPSK to OOK as CSI diminishes (or as SNR increases) results in dramatic gains in the cutoff rate versus a static (pure OOK or pure BPSK) approach.

Next, we confined our interest to BPSK and determined the optimal energy allocation between training and data for each of the three estimators under consideration. Initially, all data slots were required to have the same training energy. For the $\mathbf{Q}_{(\mathbf{1},\mathbf{0})}$ and $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimators, exact expressions for the optimal training energy were given (implicitly for the $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimator). For the $\mathbf{Q}_{(\mathbf{1},\mathbf{1})}$ estimator, analytic bounds were given at high SNR, and were supplemented by numerical simulations. For each of the three estimators, we gave a lower bound on the optimal training period that is exact at high SNR, and gave several relations that describe the relative lengths of the training period amongst the three estimators. We have shown that optimizing training over the energy allocation and training period is worthwhile; gains of up to 4 dB are to be had versus an un-optimized (but quite reasonable) training approach.

We also considered energy allocation when each data slot may have a different amount of energy allocated to it, resulting in a T-dimensional energy allocation problem. We gave an analytic solution for the optimal energy vector which is exact in several limiting senses. We have shown that, if a reasonable (close to optimal) value of T can be chosen at the transmitter, then there is little gain in a T-dimensional energy allocation in place of the simpler $2-D$ energy allocation. However, if a "loose" value of T is chosen, the T-dimensional energy allocation results in significant gain to the cutoff rate. Finally, we tested the validity of our design paradigms for energy allocation, training period, and cutoff rate for the AR(1) model, by using these same design rules when the channel correlation is described by the popular Jakes correlation model. We find that the cutoff rate predicted by the AR(1) model is extremely close the corresponding cutoff rate for the Jakes model when an equivalent value of the Doppler spread is used in each model. Furthermore, we find that the energy allocation dictated by the AR(1) model can be used in the Jakes model with little loss in the cutoff rate. Although we have focused on the AR(1) and Jakes models, our results (see section 8) hold for general channel correlation models, included the perfectly bandlimited fading channel.

There are many avenues of future work: Using average capacity as a metric, and assuming a perfectly bandlimited Doppler spectrum, optimal training frequency and energy allocation were considered in *[31]*. Extensions of our results to (possibly redundantly) precoded transmissions are of interest. Given the wide-sense stationarity of the channel, periodic placement of pilots is reasonable, but what is the optimal periodic pilot scheme? In the context of minimizing the maximum mean square estimation error, single pilot placement has been shown to be optimal for the Gauss-Markov model *[10]*. However, it is not clear that the same result holds for the cutoff rate, for it is precisely the data slot with

the largest mean square error that contributes *least* to the cutoff rate. This motivates the study of other periodic placement schemes. Next, it is expected that for fast-fading channels, superimposed training will outperform a periodic placement approach (a MMSE and BER analysis was given in *[11]* for the AR(1) model). A characterization of the region in which a form of superimposed training outperforms periodic placement is an interesting topic of further research. Also, the results in this report have been primarily for binary signaling (motivated by the popularity of binary signaling techniques, particularly at low SNR and for low complexity systems). Still, it is of interest to determine how the cutoff rate is affected by higher order constellations. Indeed, the cutoff rate expressions provided in this paper can easily be extended to arbitrary signal constellations, albeit, a closed form expression rarely exists for constellations of order larger than two.

Other extensions of interest include: a generalization of the cutoff rate analysis for the MIMO case when the receiver has partial CSI, an analysis of the merits of PSAM when on-off keying is used (i.e., is PSAM-type training beneficial in this scenario? A partial answer was given in *[26]),* a Doppler analysis when the transmitter has a statistical estimate of α (rather than a deterministic estimate, as given in section 4.4).

A. Estimator Quality Equations (Sections 2.2, 7, and 8)

Derivation of equation 5, The $Q_{(1,0)}$ estimator

The pilot observation y'_{mT} and channel state h'_{mT+k} are jointly Gaussian with correlation $\mathcal{E}[h'_{mT+k}(y'_{mT})^*] = \sqrt{E_o}\sigma_h^2 R_h(k)$. The conditional expectation of one Gaussian vector given another is evaluated in, e.g., [20]. Evaluating $\widehat{h}'_{mT+\ell} = \mathcal{E}[h'_{mT+\ell}|y'_{mT}]$ we obtain:

$$\widehat{h}'_{mT+\ell} = \frac{\sqrt{E_0}\sigma_h^2}{E_0\sigma_h^2 + \sigma_N^2} R_h(\ell) y'_{mT},$$

$$\omega_\ell^{(1,0)} = R_h^2(\ell)\frac{\kappa_0}{1+\kappa_0}.$$

Derivation of equation 6, The $Q_{(\infty,0)}$ estimator

Let $\mathbf{y} = [y'_0, y'_{-T}, \ldots, y'_{-(M-1)T}]$ be the last M pilot observations. Then, from [20, pp. 508–509], the $Q_{(M,0)}$ estimator is described by

$$\widehat{h}'_{mT+\ell} = \mathcal{E}[h'_{mT+\ell}|\mathbf{y}] = C_{h'_{mT+\ell}\mathbf{y}}C_{\mathbf{yy}}^{-1}\mathbf{y}, \text{ and} \tag{31}$$

$$\omega_\ell^{(M,0)} = \frac{1}{\sigma_h^2}C_{h'_{mT+\ell}\mathbf{y}}C_{\mathbf{yy}}^{-1}C_{h'_{mT+\ell}\mathbf{y}}^H \tag{32}$$

where

$$(C_{\mathbf{yy}})_{ij} = (\mathcal{E}[\mathbf{y}\mathbf{y}^\mathbf{H}])_{ij} = E_0\sigma_h^2 R_h(|i-j|T) + \sigma_N^2\delta(|i-j|), \text{ and}$$

$$\left(C_{h'_{mT+\ell}\mathbf{y}}\right)_{1j} = \mathcal{E}[h'_{mT+\ell}\mathbf{y}] = \sqrt{E_o}\sigma_h^2 R_h(jT+\ell).$$

Next, define the function

$$z(\tau) \triangleq \mathcal{Z}^{-1}\left\{\frac{1}{\mathcal{Z}[R_h(T\tau)] + \frac{1}{\kappa_0}}\right\}. \tag{33}$$

As $M \to \infty$, spectral factorization can be used to show that

$$\left(C_{\mathbf{yy}}^{-1}\right)_{ij} = \frac{1}{E_o\sigma_h^2}z(|i-j|). \tag{34}$$

under mild conditions on $R_h(T\tau)$. Substituting equation 34 into 31 and 32 we find that the estimate and estimator quality for the $Q_{(\infty,0)}$ estimator are given by

$$\widehat{h}'_{mT+\ell} = \frac{1}{\sqrt{E_o}}\sum_{v=0}^{\infty}y'_{vT+\ell}\gamma_{v,\ell} \text{ and} \tag{35}$$

$$\omega_\ell^{(\infty,0)} = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}R_h(nT+\ell)R_h(mT+\ell)z(|n-m|) \tag{36}$$

47

where $\gamma_{v,\ell} = \sum_{n=0}^{\infty} R_h(nT + \ell)z(|n - v|)$.

Derivation of equation 7, The $Q_{(1,1)}$ estimator

Observe that

$$
\begin{bmatrix} h'_{mT+\ell} \\ y'_{mT} \\ y'_{(m+1)T} \end{bmatrix} \sim \mathcal{CN} \left(\mathbf{0}, \begin{bmatrix} \sigma_h^2 & R_h(\ell)\sqrt{E_o}\sigma_h^2 & R_h(T-\ell)\sqrt{E_o}\sigma_h^2 \\ R_h(\ell)\sqrt{E_o}\sigma_h^2 & \sigma_h^2 E_0 + \sigma_N^2 & R_h(T)E_o\sigma_h^2 \\ R_h(T-\ell)\sqrt{E_o}\sigma_h^2 & R_h(T)E_o\sigma_h^2 & \sigma_h^2 E_0 + \sigma_N^2 \end{bmatrix} \right)
$$

Evaluating $\widehat{h}'_{mT+\ell} = \mathcal{E}\left[h'_{mT+\ell} | y'_{mT}, y'_{(m+1)T} \right]$ we obtain:

$$
\widehat{h}'_{mT+\ell} = \frac{\sigma_h^2}{\sigma_N^2} \sqrt{E_0} \left[\Gamma_{(\ell)} \, y'_{mT} + \Gamma_{(T-\ell)} \, y'_{(m+1)T} \right],
$$

$$
\omega_\ell^{(1,1)} = (\Gamma_{(\ell)}^2 + \Gamma_{(T-\ell)}^2)(\kappa_0^2 + \kappa_0) + 2\kappa_0^2 \Gamma_{(\ell)}\Gamma_{(T-\ell)}R_h(T),
$$

where

$$
\Gamma_{(k)} \triangleq \frac{R_h(k)(\kappa_0 + 1) - R_h(T)R_h(T - k)\kappa_0}{(\kappa_0 + 1)^2 - \kappa_0^2 \, R_h^2(T)}.
$$

B. The Infinite Past Estimator for the AR(1) Model (Section 2.2)

In Appendix A. we found the estimator quality for the $\mathbf{Q}_{(\infty,0)}$ estimator, for a general channel correlation $R_h(\tau)$. Alternatively, the variance has been derived in [20, pp.443] using standard Kalman filter theory. The estimator quality $\omega_\ell^{(\infty,0)}$ is given implicitly to be (using the current notation)

$$\omega_\ell^{(\infty,0)} = \frac{\frac{\sigma_N^2}{E_0}\left[\alpha^{2T}\omega_\ell^{(\infty,0)} + \left(1 - \alpha^{2T}\right)\right]}{\alpha^{2T}\sigma_h^2\omega_\ell^{(\infty,0)} + \sigma_h^2(1 - \alpha^2) + \frac{\sigma_N^2}{E_0}}. \tag{37}$$

Solving for the estimator variance yields

$$\omega_\ell^{(\infty,0)} = \alpha^{2\ell}\left[1 - \frac{1}{\frac{1}{2}\left(1 + \kappa_0\right) + \sqrt{\frac{1}{4}(1 + \kappa_0)^2 + \kappa_0\frac{\alpha^{2T}}{1 - \alpha^{2T}}}}\right].$$

Intentionally Left Blank.

C. Interleaving (Section 2.2)

The system operates on codewords of length $N' = N(T-1)$. Without loss of generality, consider the codeword that starts at time $k = 0$ denoted by

$$\mathbf{v} = \left[s_1 \ldots s_{T-1}, s_{T+1}, \ldots, s_{2T-1}, \ldots, s_{(N-1)T+1}, \ldots, s_{NT-1} \right]^t.$$

Similarly, let $\mathbf{h}' = \left[h'_1 \ldots h'_{T-1}, h'_{T+1}, \ldots, h'_{NT-1} \right]^t$ denote the channel,
$\widehat{\mathbf{h}}' = \left[\widehat{h}'_1 \ldots \widehat{h}'_{T-1}, \widehat{h}'_{T+1}, \ldots, \widehat{h}'_{NT-1} \right]^t$ denote the channel estimate, and
$\widetilde{\mathbf{h}}' = \left[\widetilde{h}'_1 \ldots \widetilde{h}'_{T-1}, \widetilde{h}'_{T+1}, \ldots, \widetilde{h}'_{NT-1} \right]^t$ denote the channel estimation error during the
transmission of a codeword. It follows that $\Sigma \triangleq \frac{1}{\sigma_h^2} \left(\mathcal{E} \left[\mathbf{h}' \mathbf{h}'^{\mathbf{H}} \right] \right)_{ij} = R_h(|i - j|)$, where $R_h(.)$
is the normalized channel correlation function. Similarly, we define normalized correlation matrices for the channel estimate and estimation error,

$$\widehat{\Sigma} \triangleq \frac{1}{\sigma_h^2} \mathcal{E} \left[\widehat{\mathbf{h}} \widehat{\mathbf{h}}^H \right] \text{ and } \widetilde{\Sigma} \triangleq \frac{1}{\sigma_h^2} \mathcal{E} \left[\widetilde{\mathbf{h}} \widetilde{\mathbf{h}}^H \right].$$

The output is $\mathbf{y}' = \sqrt{E} \widehat{\mathbf{h}}' \odot \mathbf{v} + \sqrt{E} \widetilde{\mathbf{h}}' \odot \mathbf{v} + \mathbf{n}$, where $E \triangleq \text{diag} \left\{ E_\ell \right\}_{\ell=1}^{T-1} \otimes I_N$ is the energy matrix, and where the noise vector $\mathbf{n} = \left[n_1 \ldots n_{T-1}, n_{T+1}, \ldots, n_{NT-1} \right]^t$. The matrices $\widetilde{\Sigma}$ and $\widehat{\Sigma}$ depend on the particular estimation scheme, but, in general, the diagonal elements are given by $\widehat{\Sigma} \odot I_{N'} = I_N \otimes \text{diag} \left\{ \omega_\ell \right\}_{\ell=1}^{T-1}$ and $\widetilde{\Sigma} \odot I_{N'} = I_N \otimes \text{diag} \left\{ 1 - \omega_\ell \right\}_{\ell=1}^{T-1}$.

The cutoff rate for generalized binary signaling *without interleaving* can be found in a manner similar to that described in section 3. Considering a super-symbol of length N', the cutoff rate is given by

$$R_o = \lim_{N \to \infty} \max_{Q(.)} -\frac{1}{NT} \log_2 \int_{\mathbf{y}'} \int_{\widehat{\mathbf{h}}'} \left[\sum_{\mathbf{v} \in \mathcal{S}_1^N \times \ldots \times \mathcal{S}_{T-1}^N} Q(\mathbf{v}) \sqrt{P(\mathbf{y}, \widehat{\mathbf{h}} \mid \mathbf{v})} \right]^2 d\widehat{\mathbf{h}}' \, d\mathbf{y}'.$$

Evaluating, the non-interleaved cutoff rate is

$$R_o = -\lim_{N \to \infty} \min_{Q(.)} \frac{1}{NT} \log_2 \sum_{\mathbf{v}, \mathbf{w} \in \mathcal{S}_1^N \times \ldots \times \mathcal{S}_{T-1}^N} Q(\mathbf{v}) Q(\mathbf{w})$$

$$\frac{\left| I_{N'} + \mathcal{K} V \widetilde{\Sigma} V^H \right|^{1/2} \left| I_{N'} + \mathcal{K} W \widetilde{\Sigma} W^H \right|^{1/2}}{\left| I_{N'} + \frac{1}{2} \mathcal{K} \left(V \widetilde{\Sigma} V^H + W \widetilde{\Sigma} W^H \right) + \frac{1}{4} \mathcal{K} (V - W) \widehat{\Sigma} (V - W)^H \right|}, \quad (38)$$

where $V = \text{diag} \left\{ \mathbf{v} \right\}$, $W = \text{diag} \left\{ \mathbf{w} \right\}$, and where $\mathcal{K} \triangleq \text{diag} \left\{ \kappa_\ell \right\}_{\ell=1}^{T-1} \otimes I_N$ is the *faded* energy matrix.

In practice, P-depth (finite) interleavers are often used to reduce the correlation within a transmitted codeword. The P-depth interleaver collects a codeword for transmission; it then transmits consecutive letters of the codeword once every $P-1$ symbols slots. To avoid a loss in overall transmission rate, P codewords are multiplexed in this fashion. For channels which can be accurately modelled as wide sense stationary with a monotonically decreasing correlation function (e.g., a channel described by a first order AR(1) process, $R_h(\tau) = R_A(\tau) \triangleq \alpha^{|\tau|}, \ -1 \le \alpha \le 1$), the correlation of the channel within a codeword decreases as the depth of the interleaver P increases. For channels whose correlation function can only be bounded by an exponentially decreasing envelope, i.e., $|R_h(\tau)| \le \exp^{-\psi |\tau|}, \ \psi > 0$, correlation is not reduced for each increase in P, but, as a general trend, it does decrease with increasing P. If the value of P is chosen large enough for this class of channels, the correlation within a codeword can be made arbitrarily small, presumably leading to better system performance (at the expense of added complexity, and delay) when using codes designed for independently fading channels.

A common assumption used in the analysis of correlated fading channels is that of a "perfect" (or infinite depth) interleaver at the transmitter and a corresponding deinterleaver at the receiver. The infinite depth interleaver simplifies the analysis of correlated fading channels, and is often an implicit assumption when i.i.d. channel models are used in analysis. The infinite depth interleaver effectively removes the correlation of the channel within a codeword transmission. The assumption of infinite interleaving is equivalent to setting the non-diagonal entries of the correlation matrices $\widehat{\Sigma}$ and $\widetilde{\Sigma}$ to 0, so that

$$\widehat{\Sigma} \leftarrow \widehat{\Sigma} \odot I_{N'} = I_N \otimes \text{diag}\left\{\omega_\ell\right\}_{\ell=1}^{T-1}, \quad \text{and}$$
$$\widetilde{\Sigma} \leftarrow \widetilde{\Sigma} \odot I_{N'} = I_N \otimes \text{diag}\left\{1-\omega_\ell\right\}_{\ell=1}^{T-1}.$$

Also, given the perfect interleaving assumption, $\widehat{\mathbf{h}}'$ and $\widetilde{\mathbf{h}}'$ are mutually independent, i.e., $\mathcal{E}\left[\widehat{\mathbf{h}}' \ \widetilde{\mathbf{h}}'^{H}\right] = 0_{N'}$, so that the ML-decoder, generally given by

$$\widehat{\mathbf{v}} = \max_{\mathbf{v}} P(\mathbf{y}', \widehat{\mathbf{h}}' \mid \mathbf{v})$$

reduces to a product-wise detector

$$\widehat{\mathbf{v}} = \max_{\mathbf{v}} \prod_{k=1}^{N'} P(y_k, \widehat{h}_k \mid s_k).$$

where the sequence $\left\{\widehat{h}_k\right\}$ is an independent sequence having the same marginal statistics as $\left\{\widehat{h}'_k\right\}$, as described the post-estimation post-interleaving system equation 30.

52

Note that under perfect interleaving,

$$Q(\mathbf{v}) = \left[\prod_{\ell=1}^{T-1} Q_\ell(s_k)\right]^{N(T-1)},$$

and that equation 37 reduces to the interleaved cutoff rate given by equation 10.

Intentionally Left Blank.

D. Cutoff Rate Derivation (Section 3)

Here, we evaluate equation 9, and show that it yields equation 10. Expanding equation 9

$$R_o = -\min_{Q(.)} \frac{1}{T} \log_2 \sum_{\mathbf{s},\mathbf{v} \in \mathcal{S}_1 \times \ldots \times \mathcal{S}_{T-1}} Q(\mathbf{s})Q(\mathbf{v})\mathcal{E}_{\widehat{\mathbf{h}}} \left[\int_{\mathbf{y}} \sqrt{P(\mathbf{y}|\mathbf{s},\widehat{\mathbf{h}})} \sqrt{P(\mathbf{y}|\mathbf{v},\widehat{\mathbf{h}})} \, d\mathbf{y} \right]. \qquad (39)$$

Note that $\widehat{\mathbf{h}} \sim \mathcal{CN}(0,\widehat{\Sigma})$, where $\widehat{\Sigma} = \text{diag}\{\omega_\ell\}_{\ell=1}^{T-1}$. Next, we make the following definitions: $E \triangleq \text{diag}\{E_\ell\}_{\ell=1}^{T-1}$, $\mathbf{u}_s \triangleq \sqrt{E}S\widehat{\mathbf{h}}$, $\mathbf{u}_v \triangleq \sqrt{E}V\widehat{\mathbf{h}}$, $\Sigma_s \triangleq \sigma_h^2 ES\widetilde{\Sigma}S^H + \sigma_N^2 I_{T'}$, and $\Sigma_v \triangleq \sigma_h^2 EV\widetilde{\Sigma}V^H + \sigma_N^2 I_{T'}$ where $S = \text{diag}\{\mathbf{s}\}$, $V = \text{diag}\{\mathbf{v}\}$, and $\widetilde{\Sigma} = \text{diag}\{1 - \omega_\ell\}_{\ell=0}^{T-1}$. Note that

$$\mathbf{y}|\mathbf{s},\widehat{\mathbf{h}} \sim \mathcal{CN}(\mathbf{u}_s, \Sigma_s),$$
$$\mathbf{y}|\mathbf{v},\widehat{\mathbf{h}} \sim \mathcal{CN}(\mathbf{u}_v, \Sigma_v),$$

and that

$$P(\mathbf{y}|\mathbf{s},\widehat{\mathbf{h}})P(\mathbf{y}|\mathbf{v},\widehat{\mathbf{h}})$$
$$= \frac{\exp\left\{-(\mathbf{y}-\mathbf{u}_s)^H\Sigma_s^{-1}(\mathbf{y}-\mathbf{u}_s) - (\mathbf{y}-\mathbf{u}_v)^H\Sigma_v^{-1}(\mathbf{y}-\mathbf{u}_v)\right\}}{\pi^{2T'}|\Sigma_s\Sigma_v|}$$
$$= \frac{\exp\left\{-\mathbf{y}^H(\Sigma_s^{-1} + \Sigma_v^{-1})\mathbf{y} + 2\text{Re}\left(\mathbf{y}^H(\Sigma_s^{-1}\mathbf{u}_s + \Sigma_v^{-1}\mathbf{u}_v) - \mathbf{u}_s\Sigma_s^{-1}\mathbf{u}_s - \mathbf{u}_v\Sigma_v^{-1}\mathbf{u}_v\right)\right\}}{\pi^{2T'}|\Sigma_s\Sigma_v|}.$$

Evaluating the integral, we obtain

$$\int_{\mathbf{y}} \sqrt{P(\mathbf{y}|\mathbf{s},\widehat{\mathbf{h}})P(\mathbf{y}|\mathbf{v},\widehat{\mathbf{h}})}$$
$$= \mathcal{E}_{\mathbf{z}}\left[e^{\text{Re}\left\{\ \mathbf{z}^H(\Sigma_s^{-1}\mathbf{u}_s + \Sigma_v^{-1}\mathbf{u}_v)\right\}}\right] e^{-\frac{1}{2}\left(\mathbf{u}_s^H\Sigma_s^{-1}\mathbf{u}_s + \mathbf{u}_v^H\Sigma_v^{-1}\mathbf{u}_v\right)} \frac{\left|\left(\frac{\Sigma_s^{-1} + \Sigma_v^{-1}}{2}\right)^{-1}\right|}{|\Sigma_s\Sigma_v|^{1/2}},$$

where $\mathbf{z} \sim \mathcal{CN}\left(\mathbf{0}, \left(\frac{\Sigma_s^{-1} + \Sigma_v^{-1}}{2}\right)^{-1}\right)$. Evaluating the expectation and simplifying the ratio of determinants above, we get

$$\int_{\mathbf{y}} \sqrt{P(\mathbf{y}|\mathbf{s},\widehat{\mathbf{h}})P(\mathbf{y}|\mathbf{v},\widehat{\mathbf{h}})}$$
$$= e^{\frac{1}{4}\left(\Sigma_s^{-1}\mathbf{u}_s + \Sigma_v^{-1}\mathbf{u}_v\right)^H\left(\frac{\Sigma_s^{-1} + \Sigma_v^{-1}}{2}\right)^{-1}\left(\Sigma_s^{-1}\mathbf{u}_s + \Sigma_v^{-1}\mathbf{u}_v\right)} e^{-\frac{1}{2}\left(\mathbf{u}_s^H\Sigma_s^{-1}\mathbf{u}_s + \mathbf{u}_v^H\Sigma_v^{-1}\mathbf{u}_v\right)} \frac{|\Sigma_s|^{1/2}\,|\Sigma_v|^{1/2}}{\left|\frac{\Sigma_s + \Sigma_v}{2}\right|}$$
$$= e^{-\frac{1}{2}(\mathbf{u}_s - \mathbf{u}_v)^H(\Sigma_s + \Sigma_v)^{-1}(\mathbf{u}_s - \mathbf{u}_v)} \frac{|\Sigma_s|^{1/2}\,|\Sigma_v|^{1/2}}{\left|\frac{\Sigma_s + \Sigma_v}{2}\right|}. \qquad (40)$$

Next, we take the expectation of equation 39 with respect to $\widehat{\mathbf{h}}$. We get (substituting back for $\mathbf{u}_s$ and $\mathbf{u}_v$):

$$= \mathcal{E}_{\widehat{\mathbf{h}}} \left[e^{-\frac{1}{2}\left(\sqrt{E}(S-V)\widehat{\mathbf{h}}\right)^H (\Sigma_s + \Sigma_v)^{-1} \left(\sqrt{E}(S-V)\widehat{\mathbf{h}}\right)} \right] \frac{|\Sigma_s|^{1/2} |\Sigma_v|^{1/2}}{\left|\frac{\Sigma_s + \Sigma_v}{2}\right|}$$

$$= \frac{|\Sigma_s|^{1/2} |\Sigma_v|^{1/2}}{\left| I_{T'} + \frac{1}{2} \sigma_h^2 \, \widehat{\Sigma} \, \sqrt{E}(S-V)^H (\Sigma_s + \Sigma_v)^{-1}(S-V)\sqrt{E} \right| \left|\frac{\Sigma_s + \Sigma_v}{2}\right|}$$

$$= \frac{|\Sigma_s|^{1/2} |\Sigma_v|^{1/2}}{\left| I_{T'} + \frac{1}{2} \sigma_h^2 \, (S-V)\sqrt{E} \, \widehat{\Sigma} \, \sqrt{E}(S-V)^H (\Sigma_s + \Sigma_v)^{-1} \right| \left|\frac{\Sigma_s + \Sigma_v}{2}\right|}$$

$$= \frac{|\Sigma_s|^{1/2} |\Sigma_v|^{1/2}}{\left| \frac{\Sigma_s + \Sigma_v}{2} + \frac{1}{4} \sigma_h^2 \, (S-V)\sqrt{E} \, \widehat{\Sigma} \, \sqrt{E}(S-V)^H \right|}.$$

Substituting for Σ_s and Σ_v, we get

$$\mathcal{E}_{\widehat{\mathbf{h}}} \left[\int_{\mathbf{y}} \sqrt{P(\mathbf{y}|S,\widehat{\mathbf{h}})P(\mathbf{y}|V,\widehat{\mathbf{h}})} \right]$$

$$= \frac{\left| I_{T'}\sigma_N^2 + \sigma_h^2 \sqrt{E}S\widetilde{\Sigma}S^H\sqrt{E} \right|^{1/2} \left| I_{T'}\sigma_N^2 + \sigma_h^2 \sqrt{E}V\widetilde{\Sigma}V^H\sqrt{E} \right|^{1/2}}{\left| I_{T'}\sigma_N^2 + \frac{1}{2} \sigma_h^2 \sqrt{E} \left(S\widetilde{\Sigma}S^H + V\widetilde{\Sigma}V^H \right) \sqrt{E} + \frac{1}{4} \sigma_h^2 \sqrt{E}(S-V)\widehat{\Sigma}(S-V)^H\sqrt{E} \right|} \tag{41}$$

All matrices above are diagonal. Simplifying equation 40 we obtain

$$\prod_{\ell=1}^{T'} \frac{\sqrt{1 + \kappa_\ell(1-\omega_\ell)|s_\ell|^2}\sqrt{1 + \kappa_\ell(1-\omega_\ell)|v_\ell|^2}}{1 + \frac{\kappa_\ell}{2}(1-\omega_\ell)(|s_\ell|^2 + |v_\ell|^2) + \frac{\kappa_\ell}{4}\omega_\ell|s_\ell - v_\ell|^2}.$$

Substituting this result into equation 38, we obtain

$$R_o = -\frac{1}{T} \sum_{\ell=1}^{T-1} \min_{Q_\ell(.)} \log_2$$

$$\sum_{s_\ell, v_\ell \in \{A_\ell, B_\ell\}} Q_\ell(s_\ell)Q_\ell(v_\ell) \frac{\sqrt{1 + \kappa_\ell(1-\omega_\ell)|s_\ell|^2}\sqrt{1 + \kappa_\ell(1-\omega_\ell)|v_\ell|^2}}{1 + \frac{\kappa_\ell}{2}(1-\omega_\ell)(|s_\ell|^2 + |v_\ell|^2) + \frac{\kappa_\ell}{4}\omega_\ell|s_\ell - v_\ell|^2}. \tag{42}$$

Finally, evaluation of the double sum in equation 41 yields equation 10.

E. OOK Optimality with No CSI (Section 3.1)

It is sufficient to consider the ℓth term in equation 10. Setting $\omega_\ell = 0$ and then dropping the subscript for brevity, we seek to minimize

$$p(1-p)\left\{\frac{\sqrt{1+\kappa|A|^2}\sqrt{1+\kappa|B|^2}}{1+\frac{\kappa}{2}\left(|A|^2+|B|^2\right)} - 1\right\}, \tag{43}$$

subject to the constraint set $\mathcal{J} = \{p|A|^2 + (1-p)|B|^2 = 1, 0 < p < 1, |A|^2, |B|^2 \geq 0\}$. For any fixed value of p, we will show that the minimizer of equation 42 is a form of OOK. Without loss of generality assume that $|A|^2 \geq 1$ and $|B|^2 \leq 1$, so that $|A|^2 = |B|^2 + \Delta_B$, where $\Delta_B \geq 0$ is a real number. Fixing p, we restate the minimization problem

$$\min_{\mathcal{J}'} \quad f(|B|^2, \Delta_B) \triangleq \frac{\sqrt{1+\kappa\left(|B|^2+\Delta_B\right)}\sqrt{1+\kappa|B|^2}}{1+\frac{\kappa}{2}\left(2|B|^2+\Delta_B\right)}$$

where $\mathcal{J}' = \{|B|^2 + p\,\Delta_B = 1, |B|^2 \geq 0, \Delta_B \geq 0\}$. Note that

$$\frac{\partial f}{\partial|B|^2} = \frac{\kappa\left[2+2\Delta_B\kappa+\Delta_B^2\kappa^2+|B|^2\kappa\left(2+\Delta_B\kappa\right)\right]}{\sqrt{1+|B|^2\kappa}\sqrt{1+\left(|B|^2+\Delta_B\right)\kappa}\left(2+|B|^2\kappa+\Delta_B\kappa\right)^2} \geq 0, \tag{44}$$

$$\frac{\partial f}{\partial\Delta_B} = -\kappa^2\frac{\left(|B|^2+\Delta_B\right)\sqrt{1+\kappa|B|^2}}{\left(2+|B|^2\kappa+\Delta_B\kappa\right)^2\sqrt{1+\left(|B|^2+\Delta_B\right)\kappa}} \leq 0. \tag{45}$$

Equation 43 implies that, for any fixed Δ_B, $|B|^2$ should be made as small as possible. The smallest possible value $|B|^2$ can take in $\mathcal{J}'$ is 0. Equation 44 implies that, for any fixed $|B|^2$, Δ_B should be made as large as possible. The largest possible value Δ_B can take in $\mathcal{J}'$ is $\frac{1}{p}$. Both of these objectives can be satisfied simultaneously, and therefore, *for any fixed p*, $|B|^2 = 0$ and $\Delta_B = |A|^2 = \frac{1}{p}$ minimizes equation 42. Therefore, when $\omega_\ell = 0$, an ON-OFF keying solution ($B = 0$) is always optimal. To find the optimal value of p (and hence, the optimal value of A), we substitute $|B|^2 = 0$ and $|A|^2 = \frac{1}{p}$ into equation 42, the optimal value of p is given implicitly by

$$p^* = \arg\min_{0<p<1} p\,(1-p)\left[\frac{\sqrt{1+\kappa\frac{1}{p}}}{1+\frac{\kappa}{2}\frac{1}{p}} - 1\right]$$

as $\kappa \to \infty$, $p^* \to \frac{1}{2}$, and as $\kappa \to 0$, $p^* \to 0$.

Intentionally Left Blank.

F.   Derivation of the Optimal Training Energy (Section 5.1)

Optimal Energy for the $\mathbf{Q_{(1,0)}}$ Estimator

For $T = 2$, the proof of the theorem follows easily. We prove the theorem for $T > 2$. Substituting the energy constraint into equation 12, we wish to maximize

$$R_o = -\frac{1}{T} \sum_{\ell=1}^{T-1} \log_2 \left[1 + \frac{1}{2} \, \alpha^{2\ell} \frac{\kappa_1^2(T-1) - \kappa_1 \kappa_{\text{tot}}}{-\kappa_1^2(T-1) + \kappa_1(\kappa_{\text{tot}} - T + 2) + \kappa_{\text{tot}} + 1} \right]. \qquad (46)$$

over the region $0 \leq \kappa_0 \leq \kappa_{\text{tot}}$.

The pth term $(1 \leq p \leq T-1)$ in the summation above has a maximizer $\kappa_{1,p}^*$ that is independent of p. Therefore, since each term in equation 45 is maximized by $\kappa_{1,p}^*$, the overall maximizer of equation 45, κ_1^*, is given simply by $\kappa_1^* = \kappa_{1,p}^*$. Finding $\kappa_1^* = \kappa_{1,p}^*$ is equivalent to

$$\kappa_1^* = \arg \min_{\kappa_1} \frac{\kappa_1^2(T-1) - \kappa_1 \kappa_{\text{tot}}}{-\kappa_1^2(T-1) + \kappa_1(\kappa_{\text{tot}} - T + 2) + \kappa_{\text{tot}} + 1}.$$

It follows readily that[§]

$$\kappa_1^* = \Gamma - \sqrt{\Gamma^2 - \frac{\kappa_{\text{tot}}}{T-1} \Gamma}, \quad \Gamma = \frac{\kappa_{\text{av}} T + 1}{T - 2}. \qquad (47)$$

Finally, we must verify that κ_1^* is in the valid range $\left[0, \frac{\kappa_{\text{tot}}}{T-1} \right]$. Note that, since $\Gamma \geq \frac{\kappa_{\text{tot}}}{T-1}$, equation 46 always yields a real number. Verifying $\kappa_1^* \geq 0$ is equivalent to verifying that the first term in equation 46 is larger in magnitude than the second term, which follows easily. To verify that $\kappa_1^* \leq \frac{\kappa_{\text{tot}}}{T-1}$, we must verify that

$$\Gamma - \sqrt{\Gamma^2 - \frac{\kappa_{\text{tot}}}{T-1} \Gamma} \leq \frac{\kappa_{\text{tot}}}{T-1},$$

or (expanding and cancelling), that $\frac{\kappa_{\text{tot}}}{T-1} \leq \Gamma$, which is true.

[§]The critical points of

$$\frac{a_1 x^2 + a_2 x}{a_3 x^2 + a_4 x + a_5}$$

are $(a_1 a_4 \neq a_2 a_3)$

$$\frac{-a_1 a_5 \pm \sqrt{a_1^2 a_5^2 - (a_1 a_4 - a_2 a_3) a_2 a_5}}{a_1 a_4 - a_2 a_3}$$

Optimal Energy for the $Q_{(\infty,0)}$ Estimator

We consider $T > 2$. Substituting the energy constraint into equation 12, we seek to maximize

$$R_o = -\frac{1}{T} \sum_{\ell=1}^{T-1} \log_2 \left[1 - \frac{\alpha^{2\ell}}{2} \frac{\kappa_{\text{tot}} - \kappa_0}{\kappa_{\text{tot}} - \kappa_0 + (T-1)} \right.$$
$$\left. \left(1 - \frac{1}{\frac{1}{2}(1+\kappa_0) + \sqrt{\frac{1}{4}(1+\kappa_0)^2 + \kappa_0 \frac{\alpha^{2T}}{1-\alpha^{2T}}}} \right) \right]$$

over the region $0 \leq \kappa_0 \leq \kappa_{\text{tot}}$. The maximizer of the pth term is given by

$$\max_{0 \leq \kappa_0 \leq \kappa_{\text{tot}}} \left[\frac{\kappa_{\text{tot}} - \kappa_0}{\kappa_{\text{tot}} - \kappa_0 + (T-1)} \right] \left[1 - \frac{1}{\frac{1}{2}(1+\kappa_0) + \sqrt{\frac{1}{4}(1+\kappa_0)^2 + \kappa_0 \frac{\alpha^{2T}}{1-\alpha^{2T}}}} \right],$$

which is independent of p. Therefore,

$$\kappa_0^* = \max_{0 \leq \kappa_0 \leq \kappa_{\text{tot}}} \left[\frac{\kappa_{\text{tot}} - \kappa_0}{\kappa_{\text{tot}} - \kappa_0 + (T-1)} \right] \left[1 - \frac{1}{\frac{1}{2}(1+\kappa_0) + \sqrt{\frac{1}{4}(1+\kappa_0)^2 + \kappa_0 \frac{\alpha^{2T}}{1-\alpha^{2T}}}} \right]. \quad (48)$$

At *low SNR*, we simplify the LHS and RHS above by noting that $\kappa_0 \leq \kappa_{\text{tot}} \to 0$, and rewrite the maximization problem as

$$\kappa_0^* = \max_{0 \leq \kappa_0 \leq \kappa_{\text{tot}}} \left[\frac{\kappa_{\text{tot}} - \kappa_0}{(T-1)} \right] \left[\kappa_0 \left(1 + \frac{\alpha^{2T}}{1 - \alpha^{2T}} \right) \right], \quad (49)$$

where we have used the fact that $\sqrt{1+x} \approx 1 + x/2$ for small x. The maximizer is seen to occur for $\kappa_0^* = \kappa_{\text{tot}}/2$. At *high SNR*, we first assume that as $\kappa_{\text{tot}} \to \infty$, the ratio $\kappa_0^*/\kappa_{\text{tot}}$ remains finite, so that $\kappa_0^* \to \infty$ as well. Under this assumption, equation 47 becomes

$$\kappa_0^* = \max_{0 \leq \kappa_0 \leq \kappa_{\text{tot}}} \frac{\kappa_{\text{tot}} - \kappa_0}{\kappa_{\text{tot}} - \kappa_0 + (T-1)} \left[1 - \frac{1}{1 + \kappa_0} \right],$$

from which we find that

$$\kappa_0^* = \frac{-\kappa_{\text{tot}} + \sqrt{\kappa_{\text{tot}}} \sqrt{T-1} \sqrt{T-2+\kappa_{\text{tot}}}}{T-2}.$$

Since $\kappa_{\text{tot}} \to \infty$, the expression above can be simplified, $\kappa_0^* = \kappa_{\text{tot}} \left[\frac{\sqrt{T-1}-1}{T-2} \right]$. Next, we assume that as $\kappa_{\text{tot}} \to \infty$, the ratio $\kappa_0^*/\kappa_{\text{tot}} \to 0$. In this case, equation 47 becomes

$$\kappa_0^* = \max_{0 \leq \kappa_0 \leq \kappa_{\text{tot}}} \left[1 - \frac{1}{\frac{1}{2}(1+\kappa_0) + \sqrt{\frac{1}{4}(1+\kappa_0)^2 + \kappa_0 \frac{\alpha^{2T}}{1-\alpha^{2T}}}} \right], \quad (50)$$

which produces the zero-rate result $\kappa_0^* = \kappa_{\text{tot}}$. Therefore, the first high SNR assumption is the correct one, i.e., $\kappa_0^* = \kappa_{\text{tot}} \left[\frac{\sqrt{T-1}-1}{T-2} \right]$ as $\kappa_{\text{tot}} \to \infty$.

Bounds on Optimal Energy for the $\mathbf{Q_{(1,1)}}$ Estimator

For the $(1,1)$ estimator, the cutoff rate is given by

$$R_o = -\frac{1}{T} \sum_{\ell=1}^{T-1} \log_2 \left[1 - \omega_\ell^{(1,1)} \frac{\kappa_1}{1+\kappa_1} \right] \tag{51}$$

where, rewriting equation 7, $\omega_\ell^{(1,1)} = \frac{\kappa_0^2 A_\ell + \kappa_0 B_\ell}{\kappa_0^2 C + 2\kappa_0 + 1}$ with $A_\ell = \alpha^{2\ell} + \alpha^{2(T-\ell)} - 2\alpha^{2T}$, $B_\ell = \alpha^{2\ell} + \alpha^{2(T-\ell)}$, and $C = 1 - \alpha^{2T}$.

When $\alpha \ll 1$, the estimator quality simplifies to $w_\ell^{(1,1)} = \frac{\kappa_0}{1+\kappa_0} \left[\max\left(\alpha^{2\ell}, \alpha^{2(T-\ell)} \right) \right]$ for $\ell = T/2$ or $w_\ell^{(1,1)} = 2\frac{\kappa_0}{1+\kappa_0}\alpha^{2\ell}$ for $T = \ell$, which implies that only the "closest" pilot aids in estimation. The optimization of equation 50 is equivalent to the maximization problem for the $\mathbf{Q_{(1,0)}}$ estimator, and yields $\kappa_{0,(1,1)}^* = \kappa_{0,(1,0)}^*$.

When $\alpha \approx 1$, the estimator quality simplifies to $w_\ell^{(1,1)} = \frac{2\kappa_0}{2\kappa_0+1}$. We find that

$$\kappa_{0,(1,1)}^* = \arg \max_{0 \le \kappa_0 \le \kappa_{\text{tot}}} \frac{\kappa_0}{\kappa_0 + 1/2} \frac{\kappa_{\text{tot}} - \kappa_0}{\kappa_{\text{tot}} - \kappa_0 + (T-1)}.$$

which yields

$$\kappa_{0,(1,1)}^* = -\mu + \sqrt{\mu^2 + \mu \kappa_{\text{tot}}}, \quad \text{where} \quad \mu = \frac{\kappa_{\text{tot}} + (T-1)}{2T - 3}. \tag{52}$$

Next, we consider the low and high SNR cases. Returning to equation 50, we denote the maximizer of the pth term in the expression above by $\kappa_{0,p}^*$, and note that

$$\kappa_{0,p}^* = \max_{0 \le \kappa_0 \le \kappa_{\text{tot}}} \left[\frac{\kappa_{\text{tot}} - \kappa_0}{\kappa_{\text{tot}} - \kappa_0 + T-1} \right] \left[\frac{\kappa_0^2 A_p + \kappa_0 B_p}{\kappa_0^2 C + 2\kappa_0 + 1} \right] \tag{53}$$

where we have used the total energy constraint in the left term above.

In general, explicit knowledge of $\kappa_{0,p}^*$ is not enough to determine κ_0^*. However, at *low SNR* $\kappa_{0,p}^* = \kappa_{\text{tot}}/2$ which is independent of p, and so $\kappa_0^* = \kappa_{0,p}^* = \kappa_{\text{tot}}/2$. To see this we rewrite equation 52 under the condition that $\kappa_0 \le \kappa_{\text{tot}} \to 0$, and obtain

$$\kappa_{0,p}^* = \max_{0 \le \kappa_0 \le \kappa_{\text{tot}}} \left[\frac{\kappa_{\text{tot}} - \kappa_0}{T-1} \right] \kappa_0 B_p. \tag{54}$$

From equation 53, it is easy to verify that $\kappa_0^* = \kappa_{0,p}^* = \kappa_{\text{tot}}/2$.

At *High SNR*, we return to equation 52, and multiply out terms to obtain

$$\kappa_{0,p}^* = \max_{0 \le \kappa_0 \le \kappa_{\text{tot}}} \frac{\kappa_0^3 A_p - \kappa_0^2(\kappa_{\text{tot}} A_p - B_p) - \kappa_0(\kappa_{\text{tot}} B_p)}{\kappa_0^3 C - \kappa_0^2 \left[C(\kappa_{\text{tot}} + Z) - 2 \right] - \kappa_0 \left[2(\kappa_{\text{tot}} + Z) - 1 \right] - (\kappa_{\text{tot}} - Z)} \tag{55}$$

61

where $Z \triangleq T{-}1$. Next, we note that the function

$$j(x) = \frac{a_1 x^3 + a_2 x^2 + a_3 x}{a_4 x^3 + a_5 x^2 + a_6 x + a_7}$$

has critical points given implicitly by

$$(a_1 a_5 - a_2 a_4)x^4 + (2a_1 a_6 - 2a_3 a_4)x^3 + (3a_1 a_7 + a_2 a_6 - a_3 a_5)x^2 + 2a_2 a_7 x + a_3 a_7 = 0.$$

Applying this fact to equation 54, we determine the relevant quartic for which at least one root equals $\kappa_{0,p}^*$:

$$\kappa_{0,p}^{*4} + b_1 \kappa_{0,p}^{*3} + b_2 \kappa_{0,p}^{*2} + b_3 \kappa_{0,p} + b_4 = 0$$

where

$$b_1 = 2\frac{(2A_p - B_p C)\kappa_{\text{tot}} + A_p(2Z - 1)}{B_p C + A_p(CZ - 2)},$$

$$b_2 = \frac{\kappa_{\text{tot}}^2 \left[-2A_p + B_p C\right] + \kappa_{\text{tot}}\left[-2A_p(-2 + Z) + B_p CZ\right] + 3A_p Z + B_p(2Z - 1)}{B_p C + A_p(CZ - 2)},$$

$$b_3 = 2\frac{-A_p \kappa_{\text{tot}}^2 + \kappa_{\text{tot}}(B_p - A_p Z) + B_p Z}{B_p C + A_p(CZ - 2)},$$

$$b_4 = \frac{-B_p \kappa_{\text{tot}}^2 - B_p \kappa_{\text{tot}} Z}{B_p C + A_p(CZ - 2)}.$$

Next, we invoke the high SNR assumption. The coefficients can be simplified:
$b_1 = \frac{4A_p - 2B_p C}{\Gamma}\kappa_{\text{tot}}$, $b_2 = \frac{-2A_p + B_p C}{\Gamma}\kappa_{\text{tot}}^2$, $b_3 = -\frac{2A_p}{\Gamma}\kappa_{\text{tot}}^2$, and $b_4 = -\frac{B_\ell}{\Gamma}\kappa_{\text{tot}}^2$, where $\Gamma \triangleq B_p C + A_p(CZ - 2)$. We continue to evaluate the quartic in this fashion (e.g., see [33, pp.11]). Retaining the relevant (positive) root, we eventually find that

$$\kappa_{0,p}^* = \kappa_{\text{tot}}\left[\gamma(p) + \sqrt{\gamma(p)^2 - \gamma(p)}\right], \qquad (56)$$

where

$$\gamma(p) = \frac{-2A_p + B_p C}{B_p C + A_p(CZ - 2)}$$

$$= \frac{\alpha^{4p} + \alpha^{2T} + \alpha^{4T} - 4\alpha^{2(p+T)} + \alpha^{2(2p+T)}}{-(T-2)\alpha^{4p} - (T-2)\alpha^{2T} + T\alpha^{4T} + 2(T-3)\alpha^{2(p+T)} + T\alpha^{2(2p+T)} - 2(T-1)\alpha^{2(p+2T)}},$$

where the last expression is in terms of fundamental quantities only. It can be verified that the root given in equation 55 (i.e., the maximizer of the pth term) is the only root within the range $0 \leq (.) \leq \kappa_{\text{tot}}$, and that $\kappa_{0,p}^*$ is decreasing for $p = 1, \ldots, \lceil\frac{T-1}{2}\rceil$ and increasing for $p = \lceil\frac{T-1}{2}\rceil + 1, \ldots, T{-}1$. This implies that, the overall maximizer, κ_0, satisfies

$$\left[\gamma\left(\lceil\frac{T-1}{2}\rceil\right) + \sqrt{\gamma\left(\lceil\frac{T-1}{2}\rceil\right)^2 - \gamma\left(\lceil\frac{T-1}{2}\rceil\right)}\right] \leq \frac{\kappa_0^*}{\kappa_{\text{tot}}} \leq \left[\gamma(1) + \sqrt{\gamma(1)^2 - \gamma(1)}\right]. \quad (57)$$

62

When $\alpha << 1$, we find that $\gamma(p) = -\frac{1}{T-2}$. Therefore, when $\kappa_{\text{tot}} \to \infty$ and $\alpha << 1$, equation 56 implies that $\frac{\kappa_0^*}{\kappa_{\text{tot}}} = \frac{1}{T-2}\left[\sqrt{T-1} - 1\right]$, since both sides of the bound are equal. When $\alpha \approx 1$, two application's of L'Hopital's rule shows that $\gamma(p) = \frac{2p^2 - 2pT + T^2}{2p^2 - 2pT - (T-2)T^2}$. Using this fact in equation 56 reveals that, at high SNR when $\alpha \approx 1$, a bound on κ_0^* is given by equation 56 where

$$\gamma(1) = \frac{T^2 - 2T + 2}{-T^3 + 2T^2 - 2T + 2},$$
$$\gamma\left(\lceil\frac{T-1}{2}\rceil\right) = \frac{T^2 + \mod(T,2)}{-2T^3 + 3T^2 + \mod(T,2)}.$$

The expressions for $\kappa_{0,(1,1)}^*$ at large SNR, for small and large α, can be evaluated for large T to show that $\kappa_{0,(1,1)}^* = \sqrt{T}\kappa_{\text{av}}$ in each case. Since $\kappa_{0,(1,1)}^*$ is a continuous function of α, we conclude that *at high SNR, $\kappa_{0,(1,1)}^* \approx \sqrt{T}\kappa_{\text{av}}$ with equality for large T.*

Proof that $\kappa_{0,(\infty,0)}^* \leq \kappa_{0,(1,0)}^*$

The optimal training energies can be written in implicit form as

$$\kappa_{0,(1,0)}^* = \arg\max_{0 \leq \kappa_0 \leq \kappa_{\text{tot}}} g(\kappa_0),$$

$$\kappa_{0,(\infty,0)}^* = \arg\max_{0 \leq \kappa_0 \leq \kappa_{\text{tot}}} g(\kappa_0)f(\kappa_0)$$

where

$$g(\kappa_0) \triangleq \left[\frac{\kappa_{\text{tot}}-\kappa_0}{\kappa_{\text{tot}} - \kappa_0 + (T-1)}\right]\left[\frac{\kappa_0}{1 + \kappa_0}\right], \quad \text{and}$$

$$f(\kappa_0) \triangleq g(\kappa_0)\left[\frac{1+\kappa_0}{\kappa_0}\right]\left[1 - \frac{1}{\frac{1}{2}(1+\kappa_0) + \sqrt{\frac{1}{4}(1+\kappa_0)^2 + \kappa_0\frac{\alpha^{2T}}{1-\alpha^{2T}}}}\right]. \tag{58}$$

Note that $g(x)$ has one critical point (a maximum) in the range $0 \leq x \leq \kappa_{\text{tot}}$ located at $x = \kappa_{0,(1,0)}^*$, and so is decreasing for $\kappa_{0,(1,0)}^* \leq x \leq \kappa_{\text{tot}}$. Note also that

$$p(\kappa_0) \triangleq \left[\frac{1+\kappa_0}{\kappa_0}\right]\left[1 - \frac{1}{\frac{1}{2}(1+\kappa_0) + \sqrt{\frac{1}{4}(1+\kappa_0)^2 + \kappa_0\frac{\alpha^{2T}}{1-\alpha^{2T}}}}\right]$$

is a decreasing function in the range $0 \leq \kappa_0 \leq \kappa_{\text{tot}}$. These two facts imply that $f(x)$ is a decreasing function for $\kappa_{0,(1,0)}^* \leq x \leq \kappa_{\text{tot}}$, and therefore, $\kappa_{0,(\infty,0)}^* = \arg\max_{0 \leq \kappa_0 \leq \kappa_{\text{tot}}} g(\kappa_0)f(\kappa_0) \leq \kappa_{0,(1,0)}^*$.

Proof that $\kappa_{0,(1,1)}^* \leq \kappa_{0,(1,0)}^*$

63

We would like to show that

$$\arg\min_{0\le \kappa_0 \le \kappa_{\mathrm{tot}}} \frac{1}{T} \sum_{\ell=1}^{T-1} \log_2 \omega_\ell^{(1,0)} \frac{\kappa_{\mathrm{av}}T - \kappa_0}{\kappa_{\mathrm{av}}T - \kappa_0 + T{-}1} \ge \arg\min_{0\le \kappa_0 \le \kappa_{\mathrm{tot}}} \frac{1}{T} \sum_{\ell=1}^{T-1} \log_2 \omega_\ell^{(1,1)} \frac{\kappa_{\mathrm{av}}T - \kappa_0}{\kappa_{\mathrm{av}}T - \kappa_0 + T{-}1}.$$

Each term in the left-hand sum is minimized by the same value of κ_0. Each term on the right-hand side has one critical point (a minimum) in $0 \le \kappa_0 \le \kappa_{\mathrm{tot}}$. Therefore, a sufficient (but not necessary) condition is to show that, for every $1 \le \ell \le T{-}1$,

$$\arg\max_{0\le \kappa_0 \le \kappa_{\mathrm{tot}}} \omega_\ell^{(1,0)} \frac{\kappa_{\mathrm{av}}T - \kappa_0}{1 + \kappa_{\mathrm{av}}T - \kappa_0} \ge \arg\max_{0\le \kappa_0 \le \kappa_{\mathrm{tot}}} \left[\frac{\omega_\ell^{(1,1)}}{\omega_\ell^{(1,0)}} \right] \omega_\ell^{(1,0)} \frac{\kappa_{\mathrm{av}}T - \kappa_0}{1 + \kappa_{\mathrm{av}}T - \kappa_0}.$$

This is proven by verifying that $\omega_\ell^{(1,1)}/\omega_\ell^{(1,0)}$ is a decreasing function of κ_0.

G. Derivation of the Training Period Bounds (Section 5.2)

Proof that High SNR Yields a Lower Bound on Training Period for Casual AR(1) Estimators

Here, we prove that for the AR(1) model, equation 15 gives a lower bound on the optimal training period that is exact at high SNR. Suppose that the last pilot *used* in the estimation is n pilots old. Note that $\omega_\ell = \omega_0 \alpha^{2\ell} = \omega_{-nT} \alpha^{2(nT+\ell)}$ for any casual estimator, and that at high SNR, $\omega_\ell = \alpha^{2(nT+\ell)}$. We wish to show that

$$\arg\max_T -\frac{1}{T} \sum_{\ell=1}^{T-1} \log_2 \left\{ 1 - \omega_{-nT} \frac{\kappa_1}{1+\kappa_1} \frac{\alpha^{2(nT+\ell)}}{2} \right\} \geq$$

$$\arg\max_T -\frac{1}{T} \sum_{\ell=1}^{T-1} \log_2 \left\{ 1 - \frac{\alpha^{2(nT+\ell)}}{2} \right\}.$$

where the right-hand side of the above equation is the cutoff rate as $\kappa_{\mathrm{av}} \to \infty$. Equivalently, we must show that

$$\frac{-\frac{1}{T+1} \sum_{\ell=1}^{T} \log_2 \left\{ 1 - \beta(T{+}1)c(T{+}1) \right\}}{-\frac{1}{T} \sum_{\ell=1}^{T-1} \log_2 \left\{ 1 - \beta(T)c(T) \right\}} \geq \frac{-\frac{1}{T+1} \sum_{\ell=1}^{T} \log_2 \left\{ 1 - c(T{+}1) \right\}}{-\frac{1}{T} \sum_{\ell=1}^{T-1} \log_2 \left\{ 1 - c(T) \right\}}$$

where $\beta(T) \triangleq w_{-nT} \frac{\kappa_1}{1+\kappa_1}$ is optimized over κ_0 and κ_1 for a fixed T, and where $c(T) \triangleq \alpha^{2(nT+\ell)}$. Next, note that $B(T)$ is an increasing function of T (a consequence of the fact that $\frac{\partial}{\partial \kappa_0} \omega_{-nT} \geq \frac{\partial}{\partial \kappa_0} \frac{\kappa_0}{1+\kappa_0}$). It is enough to show that

$$\frac{\sum_{\ell=1}^{T} \log_2 \left\{ 1 - \beta(T)c(T{+}1) \right\}}{\sum_{\ell=1}^{T-1} \log_2 \left\{ 1 - \beta(T)c(T) \right\}} \geq \frac{\sum_{\ell=1}^{T} \log_2 \left\{ 1 - c(T{+}1) \right\}}{\sum_{\ell=1}^{T-1} \log_2 \left\{ 1 - c(T) \right\}}$$

or that

$$\frac{\sum_{\ell=1}^{T} \log_2 \left\{ 1 - \beta(T)c(T{+}1) \right\}}{\sum_{\ell=1}^{T-1} \log_2 \left\{ 1 - \beta(T)c(T) \right\}}$$

is a decreasing function of $0 \leq \beta(T) \leq 1$ which is easily proven (noting that $C(T+1) \leq C(T)$). Therefore, the optimal training period at high SNR is a lower bound on the training period at any SNR; it is obviously attainable.

Proof that Training Period for Last Pilot Estimation is Larger than that for Infinite Past Estimation

Here we prove that the training period for the $\mathbf{Q}_{(1,0)}$ estimator is larger than that of the $\mathbf{Q}_{(\infty,0)}$ estimator $T^*_{(1,0)} \geq T^*_{(\infty,0)}$ for any channel with monotonically decreasing correlation function $R_h(\tau)$ under the equal energy assumption $\kappa_0 = \kappa_1 = \kappa_{\mathrm{av}}$. We wish to show that

$$\arg\max_T -\frac{1}{T}\sum_{\ell=1}^{T-1} \log_2\left[1 - \frac{R_h^2(\ell)}{2}\left(\frac{\kappa_{\mathrm{av}}}{1+\kappa_{\mathrm{av}}}\right)^2\right]$$
$$\geq \arg\max_T -\frac{1}{T}\sum_{\ell=1}^{T-1} \log_2\left[1 - R_h^2(\ell)\omega_0^{(\infty,0)}\left(\kappa_{\mathrm{av}},T\right)\frac{\kappa_{\mathrm{av}}}{1+\kappa_{\mathrm{av}}}\right].$$

where we emphasize that the estimator quality for the $\mathbf{Q}_{(\infty,0)}$ estimator is a function of κ_{av} and T. To show that the inequality above is satisfied, we show that

$$\frac{-\sum_{\ell=1}^{T} \log_2\left[1 - \frac{R_h^2(\ell)}{2}\left(\frac{\kappa_{\mathrm{av}}}{1+\kappa_{\mathrm{av}}}\right)^2\right]}{-\sum_{\ell=1}^{T-1} \log_2\left[1 - \frac{R_h^2(\ell)}{2}\left(\frac{\kappa_{\mathrm{av}}}{1+\kappa_{\mathrm{av}}}\right)^2\right]} \geq \frac{-\sum_{\ell=1}^{T} \log_2\left[1 - R_h^2(\ell)\omega_0^{(\infty,0)}\left(\kappa_{\mathrm{av}},T+1\right)\frac{\kappa_{\mathrm{av}}}{1+\kappa_{\mathrm{av}}}\right]}{-\sum_{\ell=1}^{T-1} \log_2\left[1 - R_h^2(\ell)\omega_0^{(\infty,0)}\left(\kappa_{\mathrm{av}},T\right)\frac{\kappa_{\mathrm{av}}}{1+\kappa_{\mathrm{av}}}\right]}$$

Note that, because $R_h(\tau)$ is monotone decreasing, $\omega_0^{(\infty,0)}\left(\kappa_{\mathrm{av}},T\right)$ is a decreasing function of T. We use this fact to increase the right-hand side of the above expression, and prove the resulting inequality

$$\frac{-\sum_{\ell=1}^{T} \log_2\left[1 - \frac{R_h^2(\ell)}{2}\left(\frac{\kappa_{\mathrm{av}}}{1+\kappa_{\mathrm{av}}}\right)^2\right]}{-\sum_{\ell=1}^{T-1} \log_2\left[1 - \frac{R_h^2(\ell)}{2}\left(\frac{\kappa_{\mathrm{av}}}{1+\kappa_{\mathrm{av}}}\right)^2\right]} \geq \frac{-\sum_{\ell=1}^{T} \log_2\left[1 - R_h^2(\ell)\omega_0^{(\infty,0)}\left(\kappa_{\mathrm{av}},T\right)\frac{\kappa_{\mathrm{av}}}{1+\kappa_{\mathrm{av}}}\right]}{-\sum_{\ell=1}^{T-1} \log_2\left[1 - R_h^2(\ell)\omega_0^{(\infty,0)}\left(\kappa_{\mathrm{av}},T\right)\frac{\kappa_{\mathrm{av}}}{1+\kappa_{\mathrm{av}}}\right]}$$

by showing that

$$\frac{-\sum_{\ell=1}^{T} \log_2\left[1 - \frac{R_h^2(\ell)}{2}\rho\right]}{-\sum_{\ell=1}^{T-1} \log_2\left[1 - \frac{R_h^2(\ell)}{2}\rho\right]} = 1 + \frac{-\log_2\left[1 - \frac{R_h^2(T)}{2}\rho\right]}{-\sum_{\ell=1}^{T-1} \log_2\left[1 - \frac{R_h^2(\ell)}{2}\rho\right]} \tag{59}$$

is a decreasing function in $0 \leq \rho \leq 1$, which is easily proven by noting again that $R_h^2(T) \geq R_h^2(\ell)$.

Proof that High SNR Yields a Lower Bound on Training Period for the Last Pilot Estimator

Here, we consider the $\mathbf{Q}_{(1,0)}$ estimator, and prove that for any channel with decreasing correlation function $R_h(\tau)$, equation 27 gives a lower bound on the optimal training period

that is exact at high SNR. The cutoff rate is

$$R_o = -\sum_{\ell=1}^{T-1} \log_2 \left\{ 1 - \frac{R_h^2(\ell)}{2} p(T) \right\}$$

where $p(T) \triangleq \frac{\kappa_0(T)}{1+\kappa_0(T)} \frac{\kappa_1(T)}{1+\kappa_1(T)}$ is optimized over κ_0 and κ_1 for a fixed value of T. Note that $p(T)$ is an increasing function of T. We wish to show that

$$\arg\max_T -\frac{1}{T} \sum_{\ell=1}^{T-1} \log_2 \left[1 - \frac{R_h^2(\ell)}{2} \frac{\kappa_0(T)}{1+\kappa_0(T)} \frac{\kappa_1(T)}{1+\kappa_1(T)} \right]$$

$$\geq \arg\max_T -\frac{1}{T} \sum_{\ell=1}^{T-1} \log_2 \left[1 - R_h^2(\ell) \right].$$

It is sufficient to show that

$$\frac{-\sum_{\ell=1}^{T} \log_2 \left[1 - \frac{R_h^2(\ell)}{2} p(T+1) \right]}{-\sum_{\ell=1}^{T-1} \log_2 \left[1 - \frac{R_h^2(\ell)}{2} p(T) \right]} \geq \frac{-\sum_{\ell=1}^{T} \log_2 \left[1 - \frac{R_h^2(\ell)}{2} \right]}{-\sum_{\ell=1}^{T-1} \log_2 \left[1 - \frac{R_h^2(\ell)}{2} \right]}$$

We can replace the left hand side of the above expression by a smaller quantity and show that the resulting inequality holds. Replacing, $p(T+1)$ with $p(T)$, we must show that

$$\frac{-\log_2 \left[1 - \frac{R_h^2(T)}{2} p(T) \right]}{-\sum_{\ell=1}^{T-1} \log_2 \left[1 - \frac{R_h^2(\ell)}{2} p(T) \right]} \geq \frac{-\log_2 \left[1 - \frac{R_h^2(T)}{2} \right]}{-\sum_{\ell=1}^{T-1} \log_2 \left[1 - \frac{R_h^2(\ell)}{2} \right]},$$

or that the left-hand side is a decreasing function in $0 \leq p(T) \leq 1$. A sufficient condition is to show that

$$f(p) = \frac{-\log_2 \left[1 - p\frac{R_h^2(\ell)}{2} \right]}{-\log_2 \left[1 - p\frac{R_h^2(T)}{2} \right]}$$

is increasing in the range $0 \leq p \leq 1$ which follows by nothing that $0 \leq R_h(T) \leq R_h(\ell) \leq 1$.

Intentionally Left Blank.

H. Variable-Energy Substitution Function (Section 6.1)

First we show that $\widetilde{\boldsymbol{\kappa}}^* = \boldsymbol{\kappa}^*$ as $\alpha \to 1$. As $\alpha \to 1$, all data slots following a pilot have the same estimator quality or "predictability," and so we expect all data slots to be activated and for $\boldsymbol{\kappa}^*$ to converge to the solution for the fixed-data energy case equation 13. That is, we expect $\boldsymbol{\kappa}^* = [\kappa_0, \kappa_1, \ldots, \kappa_1]$, where κ_0 and κ_1 are given in equation 13. This is the case, as will be shown below.

From equation 20, we activate all $T-1$ data slots iff $\kappa_{\text{tot}} \geq \phi_\alpha(T-2)$. Note that

$$\lim_{\alpha \to 1} \phi_\alpha(T-2) = (T-2) - (T - \frac{3}{2}) + \sqrt{\frac{1}{4}} = 0,$$

and, therefore, all $T-1$ data slots are activated. Note that

$$\lim_{\alpha \to 1} \kappa_\ell = \frac{1}{T-1}\left[\kappa_{\text{tot}} - \kappa_0(T-1) + (T-1)\right] - 1$$
$$= \frac{\kappa_{\text{tot}} - \kappa_0(T-1)}{T-1}, \tag{60}$$

Therefore, each data slot is allocated an equal amount of data energy. Next, note that

$$\lim_{\alpha \to 1} \kappa_0(T-1) = -\frac{\kappa_{\text{tot}}+T-1}{T-2} + \sqrt{\frac{T-1}{(T-2)^2}(\kappa_{\text{tot}}+T-1)^2 - \frac{T-1}{T-2}(\kappa_{\text{tot}}+T-1)}. \tag{61}$$

Substituting equation 60 into 59 and simplifying, we get

$$\kappa_\ell = \frac{\kappa_{\text{tot}}+1}{T-2} - \sqrt{\frac{(T-1)(\kappa_{\text{tot}}+T-1)^2 - (T-1)(T-2)(\kappa_{\text{tot}}+T-1)}{(T-2)^2(T-1)^2}}$$
$$= \frac{\kappa_{\text{tot}}+1}{T-2} - \sqrt{\frac{(\kappa_{\text{tot}}+1)(\kappa_{\text{tot}}+T-1)}{(T-1)(T-2)^2}}$$

which, with minor simplification, is seen to match equation 13.

Here, we show that $\widetilde{\boldsymbol{\kappa}}^* = \boldsymbol{\kappa}^*$ as $\alpha \to 0$. As $\alpha \to 0$, predictability of the channel is lost and we expect only one data channel to be activated, and allocated half of the total available energy (for $T_A = 1$, the variable data energy problem reduces to the fixed data energy problem of equation 13). Next, we show that this is the case: From equation 20, we activate only one data channel if $0 \leq \kappa_{\text{tot}} < \phi_\alpha(2)$. Since $\lim_{\alpha \to 0} \phi_\alpha(2) = \infty$, only one data channel should be activated. From part (c) of equation 20, we see that the activated data slot if allocated half the total energy.

Here, we show that $\widetilde{\boldsymbol{\kappa}}^* = \boldsymbol{\kappa}^*$ as $\kappa_{\text{tot}} \to 0$. Note that $\kappa_{\text{tot}} \to 0$ implies that $\kappa_\ell \to 0$, $0 \leq \ell \leq T-1$. We start from the cutoff rate expression in equation 18 and note

that, since $\alpha^{2\ell} \leq 1$,

$$R_o = -\frac{1}{T} \sum_{\ell=1}^{T-1} \frac{1}{2} \alpha^{2\ell} \frac{\kappa_0}{1+\kappa_0} \frac{\kappa_\ell}{1+\kappa_\ell} + O\left[\frac{1}{4} \alpha^{4\ell} \left(\frac{\kappa_0}{1+\kappa_0}\right)^2 \left(\frac{\kappa_\ell}{1+\kappa_\ell}\right)^2\right]$$

where a Taylor series expansion has been taken (around 0) for each term in the sum of equation 18. As $\kappa_{\text{tot}} \to 0$

$$R_o = -\frac{1}{T} \sum_{\ell=1}^{T-1} \frac{1}{2} \alpha^{2\ell} \frac{\kappa_0}{1+\kappa_0} \frac{\kappa_\ell}{1+\kappa_\ell} \tag{62}$$

It is clear that optimization of equation 61 and 19 over $\boldsymbol{\kappa}$ yields the same optimizer, and therefore, that $\widetilde{\boldsymbol{\kappa}}^* = \boldsymbol{\kappa}^*$ as $\alpha \to 0$.

I. Energy Allocation for Variable-Energy Data Slots (Section 6.2)

We will give the proof of the theorem for the case where $T_A > 1$. First, we verify that $\widetilde{R}_0$ is concave in $\boldsymbol{\kappa}$. This follows easily:

$$\frac{\partial^2 \widetilde{R}_o}{\partial \kappa_0^2} = -\frac{2}{(1+\kappa_0)^3} \sum_{p=1}^{T-1} \alpha^{2p} \frac{\kappa_p}{1+\kappa_p} \leq 0,$$

$$\frac{\partial \widetilde{R}_o^2}{\partial \kappa_\ell^2} = -2\alpha^{2\ell} \frac{\kappa_0}{(1+\kappa_0)(1+\kappa_\ell)^3} \leq 0, \quad \text{for } 1 \leq \ell \leq T-1.$$

Suppose M slots are active (the optimal value of $M = T_A$ will follow from the analysis). From the Kuhn-Tucker conditions [22], a necessary and sufficient condition for the elements of $\boldsymbol{\kappa}$ to optimize $\widetilde{R}_o$ (given that M slots are active) is

$$\frac{\partial \widetilde{R}_o}{\partial \kappa_\ell} \mid_{\kappa_\ell} = \lambda, \quad \text{for } 0 \leq \ell \leq M, \tag{63}$$

$$\frac{\partial \widetilde{R}_o}{\partial \kappa_\ell} \mid_{\kappa_\ell = 0} \leq \lambda, \quad \text{for } M+1 \leq \ell \leq T-1, \tag{64}$$

Consider the following candidate solution: let the training energy be given by

$$\kappa_0(M) = -\Delta\left(M + \kappa_{\text{tot}}\right)$$
$$+ \sqrt{\left(\Delta^2 + \Delta\right)\left(M + \kappa_{\text{tot}}\right)^2 - \left(\Delta + 1\right)\left(M + \kappa_{\text{tot}}\right)}, \tag{65}$$

where $\Delta = \frac{1}{2} \frac{(1-\alpha)(1+\alpha^M)}{\alpha - \alpha^M}$, and let the data energy be given by

$$\kappa_\ell = \begin{cases} \alpha^{\ell-1} \frac{1-\alpha}{1-\alpha^M}\left[\kappa_{\text{tot}} - \kappa_0(M) + M\right] - 1, & \text{for } 1 \leq \ell \leq M \\ 0, & \text{for } M+1 \leq \ell \leq T-1. \end{cases} \tag{66}$$

Substituting equations 64 and 65 into equation 62, we see that the condition is satisfied and that

$$\lambda = \alpha^2 \frac{\kappa_0}{1+\kappa_0}\left(\frac{1-\alpha^M}{1-\alpha}\right)^2 \frac{1}{\left(\kappa_{\text{tot}} - \kappa_0 + M\right)^2},$$

(we abbreviate $\kappa_0(M)$ as κ_0 for brevity). Substituting equation 65 into 63, we see that the condition in equation 63 is satisfied only if

$$\alpha^{2(M+1)} \frac{\kappa_0}{1+\kappa_0} \leq \lambda = \alpha^2 \frac{\kappa_0}{1+\kappa_0}\left(\frac{1-\alpha^M}{1-\alpha}\right)^2 \frac{1}{\left(\kappa_{\text{tot}} - \kappa_0 + M\right)^2},$$

or, equivalently, only if

$$\kappa_{\text{tot}} - \kappa_0 + M \leq \frac{\alpha^{-M} - 1}{1 - \alpha} \tag{67}$$

71

Substituting equation 64 into 66 and solving explicitly for κ_0, we see that this condition is equivalently satisfied if

$$\kappa_{\text{tot}} \leq \left[\frac{\alpha^{-M} - 1}{1 - \alpha} - \left(M + \frac{1}{2} \right) + \sqrt{\frac{1}{4} + \frac{\left(\alpha^{-M} - \alpha \right)\left(\alpha^{-M} - 1 \right)}{1 - \alpha^2}} \right], \qquad (68)$$

Therefore, if *the initial* choice of M satisfies equation 67, then equations 64 and 65 are indeed the maximizers of $\widetilde{R}_o$. Rewording: first we choose T_A to be the smallest integer in $\{1, \ldots, T-1\}$ such that equation 67 holds (or, if no such integer exists, choose $T_A = T - 1$). Then, we select the training and data energies as in equations 67 and 64. This establishes equation 20.

References

[1] Abou-Faycal, I.; Trott, M.; Shamai. S. The Capacity of Discrete-Time Memoryless Rayleigh-Fading Channels. *IEEE Trans. Info. Theory* **May 2001**, *47* (4), 1290–1301.

[2] Adireddy, S.; Tong, L.; Viswanathan, H. Optimal Placement of Training for Frequency Selective Block-Fading Channels. *IEEE Trans. Info. Theory* **August 2002**, *48* (8), 2338–2353.

[3] Arikan, E. An Upper Bound on the Cutoff Rate of Sequential Decoding. *IEEE Trans. Info. Theory* **Janurary 1988**, *34* (1), 55–63.

[4] Baltersee, J.; Fock, G.; Meyr, H. An Information Theoretic Foundation of Synchronized Detection. *IEEE Trans. Comm.* **December 2001**, *49* (12), 2115–2123.

[5] Berrou, C. The Ten-Year-Old Turbo Codes are Entering into Service. *IEEE Communications Magazine* **August 2003**, *41* (8), 110–116.

[6] Biglieri, E.; Proakis, J.; Shamai, S. Fading Channels: Information-Theoretic and Communications Aspects. *IEEE Trans. Info. Theory* **October 1998**, *44* (6), 2619–2692.

[7] Caire, G.; Shamai, S. On the Capacity of Some Channels with Channel Side Information. *IEEE Trans. Info Theory* **September 1999**, *45* (6), 2007–2019.

[8] Cavers, J. K. An Analysis of Pilot Symbol Assisted Modulation for Rayleigh Fading Channels [Mobile Radio]. *IEEE Trans. Veh. Tech.* **November 1991**, *40* (4), 686–693.

[9] Chen, R.-R.; Hajek, B.; Koetter, R.; Madhow U. On fixed input distributions for noncoherent communication over high SNR Rayleigh fading channels. Available at http://tesla.csl.uiuc.edu/ koetter/publications.html, June 2002.

[10] Dong, M.; Tong, L.; Sadler, B. Training placement for tracking fading channels. *Proc. IEEE ICASSP*, May 2002.

[11] Dong, M.; Tong, L.; Sadler, B. Optimal Pilot Placement for Time-Varying Channels. *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, Rome, June 2003.

[12] Gallager, R. *Information Theory and Reliable Communication*; John Wiley, 1968.

[13] Goldsmith, A.; Varaiya, P. Capacity of Fading Channels with Channel Side Information. *IEEE Info. Theory* **November 1997**, *43* (6), 1986–1992.

[14] Hassibi B.; Hochwald, B. How Much Training is Needed in Multiple-Antenna Wireless Links? *IEEE Trans. Info. Theory* **April 2003**, *49* (4), 951–963.

[15] Hero, A. O.; Marzetta, T. L. Cutoff rate and signal design for the quasi-static Rayleigh fading space-time channel. *IEEE Trans. Info. Theory* **September 2001**, *47* (6), 2400–2416.

[16] Iltis, R. A. Joint estimation of PN code delay and multipath using the extended Kalman filter. *IEEE Trans. Comm.* **October 1990**, *38* (10), 1677–1685.

[17] Jamali, S.; Le-Ngoc,T. *Coded-Modulation Techniques for Fading Channels*. Kluwer Academic Publishers, April 1994.

[18] Jakes, W. C. Jr. *Microwave Mobile Communication*. New York: Wiley, 1974.

[19] Kaplan G.; Shamai, S. (Shitz). Error probabilities for the block-fading Gaussian channel. *Arch. Electronik Übertragungstech* **1995**, *49* (4), 192–205.

[20] Kay, S. *Fundamentals of Statistical Signal Processing*. Prentice Hall, 1993.

[21] Leeuwin-Boulle, K.; Belfiore J. C. The cutoff rate of time-correlated fading channels. *IEEE Trans. Info. Theory* **March 1993**, *39* (2), 612–617.

[22] Luenberger, D. *Optimization by Vector Space Methods*, New York: Wiley, 1969.

[23] Massey, J. Coding and Moulation in Digital Communications. In *Proc. 1974 Int. Zurich Seminar., Digital Communication*, March 1974.

[24] McEliece, R.; Stark, W. Channels with Block Inerference. *IEEE Trans. Info. Theory* **Janurary 1984**, *30* (1), 44–53.

[25] Stojanovic, M.; Proakis, J. G.; Catipovic, J. A. Analysis of the impact of channel estimation errors on the performance of a decision-feedback equalizer in fading multipath channels. *IEEE Trans. Comm.* **February/March/April 1995**, *43* (2), 877–886.

[26] Medard, M.; Abou-Faycal, I.; Madhow, U. Adaptive coded modulation without channel feedback for pilot symbol assisted modulation. *Proc. 38th Annual Allerton Conference on Communication, Control, and Computing*, October 2002.

[27] Medard, M. The Effect upon Channel Capacity in Wireless Communications of Perfect and Imperfect Knowledge of the Channel. *IEEE Trans. Info. Theory* **May 2000**, *46* (3), 933–946.

[28] Misra, S.; Swami, A.; Tong, L. Cutoff Rate Analysis of the Gauss-Markov Fading Channel with Binary Inputs and Partial CSI at the Receiver. *Proc. 37th Annual Conference on Information Sciences and Systems*, March 2003.

[29] Misra, S.; Swami, A.; Tong, L. Cutoff Rate of the Gauss-Markov channel with Adaptive Energy Allocation. *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, Rome, June 2003.

[30] Misra, S.; Swami, A.; Tong, L. Optimal Training Over The Gauss-Markov Fading Channel: A Cutoff-Rate Analysis. *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montreal, May 2004.

[31] Ohno S.; Giannakis, G. Average-Rate Optimal PSAM Transmissions Over Time-Selective Fading Channels. *IEEE Trans. Wireless Comm.* **October 2002**, *1* (4), 712–720.

[32] Rappaport, T. S. *Wireless Communications Principles and Practice*, Prentice Hall, 1999.

[33] Spiegel M.; Liu, J. *Mathematical Handbook of Formulas and Tables*, McGraw Hill, 2nd edition.

[34] Wozencraft J. M.; Jacobs, I. M. *Principles of Communication Engineering*, New York, Wiley, 1965.

[35] Wu P.H.-Y.; Duel-Hallen, A. Multiuser detectors with disjoint Kalman channel estimators for synchronous CDMA mobile radio channels. *IEEE Trans. Comm.* **May 2000**, *48* (5), 732–736.

Intentionally Left Blank

Distribution List

Admnstr
Defns Techl Info Ctr
ATTN DTIC-OCP (Electrionic copy)
8725 John J Kingman Rd Ste 0944
FT Belvoir VA 22060-6218

DARPA
ATTN IXO S Welby
3701 N Fairfax Dr
Arlington VA 22203-1714

Ofc of the Secy of Defns
ATTN ODDRE (R&AT)
The Pentagon
Washington DC 20301-3080

US Army TRADOC
Battle Lab Integration & Techl Dirctrt
ATTN ATCD-B
10 Whistler Lane
FT Monroe VA 23651-5850

Dir for MANPRINT Ofc of the
Deputy Chief of Staff for Prsnnl
ATTN J Hiller
The Pentagon Rm 2C733
Washington DC 20301-0300

US Military Acdmy
Mathematical Sci Ctr of Excellence
ATTN LTC T Rugenstein
Thayer Hall Rm 226C
West Point NY 10996-1786

SMC/GPA
2420 Vela Way Ste 1866
El Segundo CA 90245-4659

US Army ARDEC
ATTN AMSTA-AR-TD
Bldg 1
Picatinny Arsenal NJ 07806-5000

US Army Avn & Mis Cmnd
ATTN AMSMI-RD W C McCorkle
Redstone Arsenal AL 35898-5240

US Army Info Sys Engrg Cmnd
ATTN AMSEL-IE-TD F Jenia
FT Huachuca AZ 85613-5300

US Army Natick RDEC
Acting Techl Dir
ATTN SBCN-TP P Brandler
Kansas Street Bldg78
Natick MA 01760-5056

US Army Simulation Train &
Instrmntn Cmnd
ATTN AMSTI-CG M Macedonia
12350 Research Parkway
Orlando FL 32826-3726

US Army Tank-Automtv Cmnd RDEC
ATTN AMSTA-TR J Chapin
Warren MI 48397-5000

Hicks & Assoc Inc
ATTN G Singley III
1710 Goodrich Dr Ste 1300
McLean VA 22102

Director
US Army Rsrch Lab
ATTN AMSRD-ARL-RO-D JCI Chang
ATTN AMSRD-ARL-RO-EN W D Bach
PO Box 12211
Research Triangle Park NC 27709

US Army Rsrch Lab
ATTN AMSRD-ARL-D J M Miller
ATTN AMSRD-ARL-CI-IS Mail & Records
Mgmt
ATTN AMSRD-ARL-CI-OK-T Techl Pub
(2 copies)
ATTN AMSRD-ARL-CI-OK-TL Techl Lib
(2 copies)
ATTN AMSRD-ARL-CI-CN S Misra
(3 copies)
Adelphi, MD 20783-1197